Polar codes for Arbitrary DMCs

Rajai Nasser and Emre Telatar, *Fellow, IEEE*
School of Computer and Communication Sciences, EPFL
Lausanne, Switzerland
Email: {rajai.nasser, emre.telatar}@epfl.ch

***Abstract*—Polar codes are constructed for arbitrary channels by imposing an arbitrary quasigroup structure on the input alphabet. Just as with "usual" polar codes, the block error probability under successive cancellation decoding is $o(2^{-N^{1/2-\epsilon}})$, where N is the block length. The encoding and decoding in this coding scheme can be implemented with a complexity of $O(N \log N)$. It is shown that the same technique can be used to construct polar codes for arbitrary multiple access channels (MAC) by using an appropriate Abelian group structure. Although the symmetric sum capacity is achieved by this coding scheme, some points in the symmetric capacity region may not be achieved. In the case where the channel is a combination of linear channels, we provide a necessary and sufficient condition characterizing the channels whose symmetric capacity region is preserved upon the polarization process. We also provide a sufficient condition for having a maximal loss in the dominant face.**

I. Introduction

Polar coding, invented by Arıkan [1], is the first low complexity coding technique that achieves the symmetric capacity of binary-input memoryless channels. Polar codes rely on a phenomenon called *polarization*, which is the process of converting a set of identical copies of a given single user binary-input channel, into a set of "almost extremal channels", i.e., either "almost perfect channels", or "almost useless channels". The probability of error of successive cancellation decoding of polar codes was proven to be equal to $o(2^{-N^{1/2-\epsilon}})$ by Arıkan and Telatar [2].

Arıkan's technique was generalized by Şaşoğlu et al. for channels with an input alphabet of prime size [3]. Generalization to channels with arbitrary input alphabet size is not simple since it was shown in [3] that if we use any group operation for the polarization method, it is not guaranteed that polarization will happen as usual to "almost perfect channels" or "almost useless channels". Şaşoğlu [4] used a special type of quasigroup operations to ensure polarization.

Park and Barg [5] showed that polar codes can be constructed using the group structure $\mathbb{Z}_{2^r}$. Sahebi and Pradhan [6] showed that polar codes can be constructed using any Abelian group structure. The polarization phenomenon described in [5] and [6] does not happen in the usual sense, indeed, it was previously proven by Şaşoğlu et al. that it is not the case. It is shown in [5] and [6] that while it is true that we don't always have polarization to "almost perfect channels" or "almost useless channels" if a general Abelian operation is used, we always have polarization to "almost useful channels" (i.e., channels that are easy to be used for communication). [5] and [6] rely mainly on the properties of Battacharyya parameters to derive polarization results. In this paper, we adopt a different approach: we give a direct elementary proof of polarization for the more general case of quasigroups using only information theoretic concepts (namely, entropies and mutual information). The Battacharyya parameter is used only to derive the rate of polarization.

In the case of multiple access channels (MAC), we find two main results in the literature: (i) Şaşoğlu et al. constructed polar codes for the two-user MAC with an input alphabet of prime size [7], (ii) Abbe and Telatar used matroid theory to construct polar codes for the m-user MAC with binary input [8]. The generalization of the results in [8] to MAC with arbitrary input alphabet size is not trivial even in the case of prime size since there is no known characterization for non-binary matroids. We have shown in [9] that the use of matroid theory is not necessary; we used elementary techniques to construct polar codes for the m-user MAC with input alphabet of prime size. In this paper, we will see how we can construct polar codes for an arbitrary MAC where the input alphabet size is allowed to be arbitrary, and possibly different from one user to another.

In our construction, as well as in both constructions in [7] and [8], the symmetric sum capacity is preserved upon the polarization process. However, a part of the symmetric capacity region may be lost in the process. In this paper, we study this loss in the special case where the channel is a combination of linear channels (this class of channels will be introduced in section 8).

In section 2, we introduce the preliminaries for this paper. We describe the polarization process in section 3. The rate of polarization is studied in section 4. Polar codes for arbitrary single user channels are constructed in section 5. The special case of group structures is discussed in section 6. We construct polar codes for arbitrary MAC in section 7. The problem of loss in the capacity region is studied in section 8. Finally, we conclude this paper in section 9.

II. Preliminaries

We first recall the definitions for multiple access channels in order to introduce the notation that will be used throughout this paper. Since ordinary channels (one transmitter and one receiver) can be seen as a special case of multiple access channels, we will not provide definitions for ordinary channels.

A part of this work was presented in the IEEE International Symposium on Information Theory, Istanbul, Turkey, July 2013.

A. Multiple access channels

Definition 1. *A discrete m-user multiple access channel (MAC) is an $(m+2)$-tuple $P = (\mathcal{X}_1,\ \mathcal{X}_2,\ ...,\ \mathcal{X}_m,\ \mathcal{Y},\ f_P)$ where $\mathcal{X}_1,\ ...,\ \mathcal{X}_m$ are finite sets that are called the* input alphabets *of P, $\mathcal{Y}$ is a finite set that is called the* output alphabet *of P, and $f_P : \mathcal{X}_1 \times \mathcal{X}_2 \times ... \times \mathcal{X}_m \times \mathcal{Y} \to [0,1]$ is a function satisfying $\forall (x_1, x_2, ..., x_m) \in \mathcal{X}_1 \times \mathcal{X}_2 \times ... \times \mathcal{X}_m$, $\sum_{y \in \mathcal{Y}} f_P(x_1, x_2, ..., x_m, y) = 1$.*

Notation 1. *We write $P : \mathcal{X}_1 \times \mathcal{X}_2 \times ... \times \mathcal{X}_m \to \mathcal{Y}$ to denote that P has m users, $\mathcal{X}_1,\ \mathcal{X}_2,\ ...,\ \mathcal{X}_m$ as input alphabets, and $\mathcal{Y}$ as output alphabet. We denote $f_P(x_1, x_2, ..., x_m, y)$ by $P(y|x_1, x_2, ..., x_m)$ which is interpreted as the conditional probability of receiving y at the output, given that $(x_1, x_2, ..., x_m)$ is the input.*

Definition 2. *A code $\mathcal{C}$ of block length N and rate vector $(R_1, R_2, ..., R_m)$ is an $(m+1)$-tuple $\mathcal{C} = (f_1, f_2, ..., f_m, g)$, where $f_k : \mathcal{W}_k = \{1, 2, ..., e^{NR_k}\} \to \mathcal{X}_k^N$ is the encoding function of the k^{th} user and $g : \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2 \times ... \times \mathcal{W}_m$ is the decoding function. We denote $f_k(w) = \big(f_k(w)_1, ..., f_k(w)_N\big)$, where $f_k(w)_n$ is the n^{th} component of $f_k(w)$. The average probability of error of the code $\mathcal{C}$ is given by:*

$$P_e(\mathcal{C}) = \sum_{(w_1,...,w_m) \in \mathcal{W}_1 \times ... \times \mathcal{W}_m} \frac{P_e(w_1, ..., w_m)}{|\mathcal{W}_1| \times ... \times |\mathcal{W}_m|},$$

$$P_e(w_1, ..., w_m) = \sum_{\substack{(y_1,...,y_N) \in \mathcal{Y}^N \\ g(y_1,...,y_N) \neq (w_1,...,w_m)}} \prod_{n=1}^{N} P\big(y_n | f_1(w_1)_n, ..., f_m(w_m)_n\big).$$

Note that e^{NR_k} has to be an integer for all $1 \leq k \leq m$.

Definition 3. *A rate vector $R = (R_1, ..., R_m)$ is said to be achievable if there exists a sequence of codes $\mathcal{C}_N$ of rate vector $(R_1 - \epsilon_{1,N}, R_2 - \epsilon_{2,N}, ..., R_m - \epsilon_{m,N})$ and of block length N such that the sequence $\{P_e(\mathcal{C}_N)\}_N$ and the sequences $\{\epsilon_{i,N}\}_N$ (for all $1 \leq i \leq m$) tend to zero as N tends to infinity. The capacity region of the MAC P is the set of all achievable rate vectors.*

Definition 4. *The information theoretic capacity region of a MAC P for input distributions $X_1, ..., X_m$ is the polymatroid region in $\mathbb{R}^m$ defined by:*

$$\mathcal{J}_{X_1,...,X_m}(P) := \big\{R = (R_1, ..., R_m) \in \mathbb{R}^m : \\ 0 \leq R(S) \leq I_{X_1,...,X_m}[S](P) \ \text{ for all } \ S \subset \{1, ..., m\}\big\}$$

where $X_1, ..., X_m$ are some independent random variables in $\mathcal{X}_1, ..., \mathcal{X}_m$ respectively. $R(S) := \sum_{k=1}^{l_S} R_k$, $X(S) := (X_{s_1}, ..., X_{s_{l_S}})$ for $S = \{s_1, ..., s_{l_S}\}$ and $I_{X_1,...,X_m}[S](P) := I(X(S); YX(S^c))$. The mutual information is computed for the probability distribution $P(y|x_1, ..., x_m)P_{X_1}(x_1)...P_{X_m}(x_m)$ on $\mathcal{X}_1 \times ... \times \mathcal{X}_m \times \mathcal{Y}$.

Theorem 1. *(Theorem 15.3.6 [10]) The capacity region of a MAC P is given by the closure of the convex hull of the union of all information theoretic capacity regions of P for all the input distributions, i.e,* $\overline{\text{ConvexHull}}\Big(\bigcup_{\substack{X_1,...,X_m \\ \text{are independent} \\ \text{random variables in} \\ \mathcal{X}_1,...,\mathcal{X}_m \text{ resp.}}} \mathcal{J}_{X_1,...,X_m}(P) \Big)$.

Definition 5. *$I_{X_1,...,X_m}(P) := I_{X_1,...,X_m}[\{1, ..., m\}](P)$ is called the* sum capacity *of P for the input distributions $X_1, ..., X_m$. It is equal to the maximum value of $R_1 + ... + R_m$ when $(R_1, ..., R_m)$ belongs to the information theoretic capacity region for input distributions $X_1, ..., X_m$. The set of points of the information theoretic capacity region satisfying $R_1 + ... + R_m = I_{X_1,...,X_m}(P)$ is called* the dominant face *of this region.*

Notation 2. *When $X_1, ..., X_m$ are independent and uniform random variables in $\mathcal{X}_1, ..., \mathcal{X}_m$ respectively, we will simply denote $\mathcal{J}_{X_1,...,X_m}(P)$, $I_{X_1,...,X_m}[S](P)$ and $I_{X_1,...,X_m}(P)$ by $\mathcal{J}(P)$, $I[S](P)$ and $I(P)$ respectively. $\mathcal{J}(P)$ is called the* symmetric capacity region *of P, and $I(P)$ is called the* symmetric sum capacity *of P.*

B. Quasigroups

Definition 6. *A quasigroup is a pair $(Q, *)$, where $*$ is a binary operation on the set Q satisfying the following:*

- *For any two elements $a, b \in Q$, there exists a unique element $c \in Q$ such that $a = b * c$. We denote this element c by $b \backslash_* a$.*
- *For any two elements $a, b \in Q$, there exists a unique element $d \in Q$ such that $a = d * b$. We denote this element d by $a /^* b$.*

Remark 1. *If $(Q, *)$ is a quasigroup, then $(Q, /^*)$ and $(Q, \backslash_*)$ are also quasigroups.*

Notation 3. *Let A and B be two subsets of a quasigroup $(Q, *)$. We define the set:*

$$A * B := \{a * b :\ a \in A, b \in B\}.$$

If A and B are non-empty, then $|A * B| \geq \max\{|A|, |B|\}$.

Definition 7. *Let Q be any set. A partition $\mathcal{H}$ of Q is said to be a* balanced partition *if and only if all the elements of $\mathcal{H}$ have the same size. We denote the common size of its elements by $||\mathcal{H}||$. The number of elements in $\mathcal{H}$ is denoted by $|\mathcal{H}|$ as usual. Clearly, $|Q| = |\mathcal{H}| \times ||\mathcal{H}||$ for such a partition.*

Definition 8. *Let $\mathcal{H}$ be a balanced partition of a set Q. We define the* projection onto $\mathcal{H}$ *as the mapping $\text{Proj}_{\mathcal{H}} : Q \longrightarrow \mathcal{H}$, where $\text{Proj}_{\mathcal{H}}(x)$ is the unique element $H \in \mathcal{H}$ such that $x \in H$.*

Lemma 1. *Let $\mathcal{H}$ be a balanced partition of a quasigroup $(Q, *)$. Define:*

$$\mathcal{H}^* := \{A * B :\ A, B \in \mathcal{H}\}.$$

If $\mathcal{H}^*$ is a balanced partition, then $||\mathcal{H}^*|| \geq ||\mathcal{H}||$.

Proof: Let $A, B \in \mathcal{H}$ then $A * B \in \mathcal{H}^*$, we have:

$$||\mathcal{H}^*|| = |A * B| \geq \max\{|A|, |B|\} = ||\mathcal{H}||.$$

■

Definition 9. *Let $(Q, *)$ be a quasigroup. A balanced partition $\mathcal{H}$ of Q is said to be a* stable partition *of $(Q, *)$ if and only if there exist n different balanced partitions $\mathcal{H}_1$, ..., $\mathcal{H}_n$ of Q such that:*

- $\mathcal{H}_1 = \mathcal{H}$.
- $\mathcal{H}_{i+1} = \mathcal{H}_i^* = \{A * B : \ A, B \in \mathcal{H}_i\}$ *for all* $i \leq n-1$.
- $\mathcal{H} = \mathcal{H}_n^*$.

n is called the degree *of $\mathcal{H}$. It is easy to see that if $\mathcal{H}$ is a stable partition of degree n, then $||\mathcal{H}_i|| = ||\mathcal{H}||$ for all $1 \leq i \leq n$ (since $\mathcal{H}_{i+1} = \mathcal{H}_i^*$ for $1 \leq i \leq n-1$, and since $\mathcal{H}_n^* = \mathcal{H}$, lemma 1 implies that $||\mathcal{H}|| = ||\mathcal{H}_1|| \leq ||\mathcal{H}_2|| \leq ... \leq ||\mathcal{H}_n|| \leq ||\mathcal{H}||$).*

Example 1. *Let $Q = \mathbb{Z}_n \times \mathbb{Z}_n$, define $(x_1, y_1) * (x_2, y_2) = (x_1 + y_1 + x_2 + y_2, y_1 + y_2)$. For each $j \in \mathbb{Z}_n$ and each $1 \leq i \leq n$, define $H_{i,j} = \{(j + (i-1)k, k) : \ k \in \mathbb{Z}_n\}$. Let $\mathcal{H}_i = \{H_{i,j} : \ j \in \mathbb{Z}_n\}$ for $1 \leq i \leq n$. It is easy to see that $\mathcal{H}_i^* = \mathcal{H}_{i+1}$ for $1 \leq i \leq n-1$ and $\mathcal{H}_n^* = \mathcal{H}_1$. Therefore, $\mathcal{H} := \mathcal{H}_1$ is a stable partition of $(Q, *)$ whose degree is n.*

Lemma 2. *If $\mathcal{H}$ is a stable partition and A_1 is an arbitrary element of $\mathcal{H}$, then $\mathcal{H}^* = \{A_1 * A_2 : \ A_2 \in \mathcal{H}\}$.*

Proof: We have:

$$Q = A_1 * Q = A_1 * \Big(\bigcup_{A_2 \in \mathcal{H}} A_2 \Big) = \bigcup_{A_2 \in \mathcal{H}} (A_1 * A_2).$$

Therefore, $\{A_1 * A_2 : \ A_2 \in \mathcal{H}\}$ covers Q and is a subset of $\mathcal{H}^*$ (which is a partition of Q that does not contain any empty element). We conclude that $\mathcal{H}^* = \{A_1 * A_2 : \ A_2 \in \mathcal{H}\}$. ■

Definition 10. *For any two partitions $\mathcal{H}_1$ and $\mathcal{H}_2$, we define:*

$$\mathcal{H}_1 \wedge \mathcal{H}_2 = \{A \cap B : \ A \in \mathcal{H}_1, \ B \in \mathcal{H}_2, \ A \cap B \neq \emptyset\}.$$

Lemma 3. *If $\mathcal{H}_1$ and $\mathcal{H}_2$ are stable then $\mathcal{H}_1 \wedge \mathcal{H}_2$ is also a stable partition of $(Q, *)$, and $(\mathcal{H}_1 \wedge \mathcal{H}_2)^* = \mathcal{H}_1^* \wedge \mathcal{H}_2^*$.*

Proof: Since $\mathcal{H}_1$ and $\mathcal{H}_2$ are two partitions of Q, it is easy to see that $\mathcal{H}_1 \wedge \mathcal{H}_2$ is also a partition of Q. Now let $A_1, A_2 \in \mathcal{H}_1$ and $B_1, B_2 \in \mathcal{H}_2$. If $A_1 \cap B_1 \neq \emptyset$ and $A_2 \cap B_2 \neq \emptyset$, we have:

$$(A_1 \cap B_1) * (A_2 \cap B_2) \ \subset \ (A_1 * A_2) \cap (B_1 * B_2) \in \mathcal{H}_1^* \wedge \mathcal{H}_2^*. \quad (1)$$

Let $A_1 \in \mathcal{H}_1$ and $B_1 \in \mathcal{H}_2$ be chosen such that $|A_1 \cap B_1|$ is maximal. Lemma 2 implies that $\mathcal{H}_1^* = \{A_1 * A_2 : \ A_2 \in \mathcal{H}_1\}$ and $\mathcal{H}_2^* = \{B_1 * B_2 : \ B_2 \in \mathcal{H}_1\}$. Therefore,

$$|Q| = \sum_{(A_2, B_2) \in \mathcal{H}_1 \times \mathcal{H}_2} |(A_1 * A_2) \cap (B_1 * B_2)|,$$

which implies that

$$|Q| \geq \sum_{\substack{(A_2, B_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \\ A_2 \cap B_2 \neq \emptyset}} |(A_1 * A_2) \cap (B_1 * B_2)| \quad (2)$$

$$\geq \sum_{\substack{(A_2, B_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \\ A_2 \cap B_2 \neq \emptyset}} |(A_1 \cap B_1) * (A_2 \cap B_2)|, \quad (3)$$

where (3) follows from (1). Now if $A_2 \cap B_2 \neq \emptyset$, we must have

$$|(A_1 \cap B_1) * (A_2 \cap B_2)| \geq |A_1 \cap B_1| \geq |A_2 \cap B_2|. \quad (4)$$

Therefore, We have:

$$\sum_{\substack{(A_2, B_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \\ A_2 \cap B_2 \neq \emptyset}} |(A_1 \cap B_1) * (A_2 \cap B_2)|$$

$$\geq \sum_{\substack{(A_2, B_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \\ A_2 \cap B_2 \neq \emptyset}} |A_1 \cap B_1| \geq \sum_{\substack{(A_2, B_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \\ A_2 \cap B_2 \neq \emptyset}} |A_2 \cap B_2| \quad (5)$$

Now since $\mathcal{H}_1$ and $\mathcal{H}_2$ are two partitions of Q, we must have $\sum_{\substack{(A_2, B_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \\ A_2 \cap B_2 \neq \emptyset}} |A_2 \cap B_2| = |Q|$. We conclude that all the inequalities in (2), (3), (4) and (5) are in fact equalities. Therefore, for all $A_2 \in \mathcal{H}_1$ and $B_2 \in \mathcal{H}_2$ such that $A_2 \cap B_2 \neq \emptyset$, we have $|A_2 \cap B_2| = |A_1 \cap B_1|$ (i.e., $\mathcal{H}_1 \wedge \mathcal{H}_2$ is a balanced partition), and $|(A_1 \cap B_1) * (A_2 \cap B_2)| = |(A_1 * A_2) \cap (B_1 * B_2)|$. Now (1) implies that $(A_1 \cap B_1) * (A_2 \cap B_2) = (A_1 * A_2) \cap (B_1 * B_2)$. Therefore, $(\mathcal{H}_1 \wedge \mathcal{H}_2)^* = \mathcal{H}_1^* \wedge \mathcal{H}_2^*$.

If $\mathcal{H}_1$ and $\mathcal{H}_2$ are of degrees n_1 and n_2 respectively, then $\mathcal{H}_1 \wedge \mathcal{H}_2$ is a stable partition whose degree is at most $\gcd(n_1, n_2)$. ■

III. Polarization process

In this section, we deal with ordinary channels having a quasigroup structure on the input alphabet.

Definition 11. *Let $(Q, *)$ be an arbitrary quasigroup, and let $P : Q \longrightarrow \mathcal{Y}$ be a single user channel. We define the two channels $P^- : Q \longrightarrow \mathcal{Y} \times \mathcal{Y}$ and $P^+ : Q \longrightarrow \mathcal{Y} \times \mathcal{Y} \times Q$ as follows:*

$$P^-(y_1, y_2 | u_1) = \frac{1}{|Q|} \sum_{u_2 \in Q} P(y_1 | u_1 * u_2) P(y_2 | u_2),$$

$$P^+(y_1, y_2, u_1 | u_2) = \frac{1}{|Q|} P(y_1 | u_1 * u_2) P(y_2 | u_2).$$

For any $s = (s_1, ..., s_n) \in \{-, +\}^n$, we define

$$P^s := ((P^{s_1})^{s_2} ...)^{s_n}.$$

Remark 2. *Let U_1 and U_2 be two independent random variables uniformly distributed in Q. Set $X_1 = U_1 * U_2$ and $X_2 = U_2$, then X_1 and X_2 are independent and uniform in Q since $*$ is a quasigroup operation. Let Y_1 and Y_2 be the outputs of the channel P when X_1 and X_2 are the inputs*

respectively. It is easy to see that $I(P^-) = I(U_1; Y_1, Y_2)$ *and* $I(P^+) = I(U_2; Y_1, Y_2, U_1)$*. We have:*

$$\begin{aligned} I(P^-) + I(P^+) &= I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2, U_1) \\ &= I(U_1, U_2; Y_1, Y_2) = I(X_1, X_2; Y_1, Y_2) \\ &= I(X_1; Y_1) + I(X_2; Y_2) = 2I(P). \end{aligned}$$

It is clear that

$$I(P^+) = I(U_2; Y_1, Y_2, U_1) \geq I(U_2; Y_2) = I(X_2; Y_2) = I(P).$$

We conclude that $I(P^-) \leq I(P) \leq I(P^+)$*.*

Definition 12. *Let* $\mathcal{H}$ *be a balanced partition of* $(Q, /^*)$*, we define the channel* $P[\mathcal{H}] : \mathcal{H} \longrightarrow \mathcal{Y}$ *by:*

$$P[\mathcal{H}](y|H) = \frac{1}{||\mathcal{H}||} \sum_{\substack{x \in Q \\ \text{Proj}_{\mathcal{H}}(x) = H}} P(y|x).$$

Remark 3. *If* X *is a random variable uniformly distributed in* Q *and* Y *is the output of the channel* P *when* X *is the input, then it is easy to see that* $I(P[\mathcal{H}]) = I(\text{Proj}_{\mathcal{H}}(X); Y)$*.*

Definition 13. *Let* $\{B_n\}_{n \geq 1}$ *be i.i.d. uniform random variables in* $\{-, +\}$*. We define the channel-valued process* $\{P_n\}_{n \geq 0}$ *by:*

$$\begin{aligned} P_0 &:= P, \\ P_n &:= P_{n-1}^{B_n} \ \forall n \geq 1. \end{aligned}$$

The main result of this section is that almost surely P_n becomes a channel where the output is "almost equivalent" to the projection of the input onto a stable partition of $(Q, /^*)$:

Theorem 2. *Let* $(Q, *)$ *be a quasigroup and let* $P : Q \longrightarrow \mathcal{Y}$ *be an arbitrary channel. Then for any* $\delta > 0$*, we have:*

$$\lim_{n \to \infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-, +\}^n : \ \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ \big| I(P^s) - \log |\mathcal{H}_s| \big| < \delta, \big| I(P^s[\mathcal{H}_s]) - \log |\mathcal{H}_s| \big| < \delta \Big\} \Big| = 1.$$

Remark 4. *Theorem 2 can be interpreted as follows: in a polarized channel* P^s*, we have* $I(P^s) \approx I(P^s[\mathcal{H}_s]) \approx \log |\mathcal{H}_s|$ *for a certain stable partition* $\mathcal{H}_s$ *of* $(Q, /^*)$*. Let* X_s *and* Y_s *be the channel input and output of* P^s *respectively.* $I(P^s[\mathcal{H}_s]) \approx \log |\mathcal{H}_s|$ *means that* Y_s *"almost" determines* $\text{Proj}_{\mathcal{H}_s}(X_s)$*. On the other hand,* $I(P^s) \approx I(P^s[\mathcal{H}_s])$ *means that there is "almost" no information about* X_s *other than* $\text{Proj}_{\mathcal{H}_s}(X_s)$ *which can be determined from* Y_s*.*

In order to prove theorem 2, we need several lemmas:

Lemma 4. *Let* $(Q, *)$ *be a quasigroup. If* A*,* B *and* C *are three non-empty subsets of* Q *such that* $|A| = |B| = |C| = |A * C| = |B * C|$*, then either* $A \cap B = \emptyset$ *or* $A = B$*.*

Proof: Suppose that $A \cap B \neq \emptyset$ and let $a \in A \cap B$. The fact that $|A * C| = |C|$ implies that $A * C = a * C$. Similarly, we also have $B * C = a * C$ since $a \in B$. Therefore, $(A \cup B) * C = a * C$, and so $|(A \cup B) * C| = |C| = |A|$. By noticing that $|A| \leq |A \cup B| \leq |(A \cup B) * C| = |A|$, we conclude that $|A \cup B| = |A|$, which implies that $A = B$ since $|A| = |B|$. ∎

Definition 14. *Let* Q *be a set, and let* A *be a subset of* Q*, we define the distribution* $\mathbb{I}_A$ *on* Q *as* $\mathbb{I}_A(x) = \frac{1}{|A|}$ *if* $x \in A$ *and* $\mathbb{I}_A(x) = 0$ *otherwise.*

Lemma 5. *Let* X *be a random variable on* Q*, and let* A *be a subset of* Q*. Suppose that there exist* $\delta > 0$ *and an element* $a \in A$ *such that* $|\mathrm{P}_X(x) - \mathrm{P}_X(a)| < \delta$ *for all* $x \in A$ *and* $\mathrm{P}_X(x) < \delta$ *for all* $x \notin A$*. Then* $||\mathrm{P}_X - \mathbb{I}_A||_\infty < |Q|\delta$*.*

Proof: We have:

$$\begin{aligned} \big|1 - |A| P_X(a)\big| &= \Big| \Big(\sum_{x \in Q} \mathrm{P}_X(x) \Big) - |A| \mathrm{P}_X(a) \Big| \\ &= \Big| \sum_{x \in A} \big(\mathrm{P}_X(x) - \mathrm{P}_X(a) \big) + \sum_{x \in A^c} \mathrm{P}_X(x) \Big| \\ &\leq \sum_{x \in A} \big| \mathrm{P}_X(x) - \mathrm{P}_X(a) \big| + \sum_{x \in A^c} \mathrm{P}_X(x) \\ &< (|Q| - 1)\delta. \end{aligned}$$

Therefore, $\big| \mathrm{P}_X(a) - \frac{1}{|A|} \big| < \frac{|Q|-1}{|A|}\delta \leq (|Q| - 1)\delta$. Let $x \in A$, then

$$\Big| \mathrm{P}_X(x) - \frac{1}{|A|} \Big| \leq \big| \mathrm{P}_X(x) - \mathrm{P}_X(a) \big| + \Big| \mathrm{P}_X(a) - \frac{1}{|A|} \Big| < |Q|\delta.$$

On the other hand, if $x \notin A$ we have $\mathrm{P}_X(x) < \delta \leq |Q|\delta$. Thus, $||P_X - \mathbb{I}_A||_\infty < |Q|\delta$. ∎

Definition 15. *Let* Q *and* $\mathcal{Y}$ *be two arbitrary sets. Let* $\mathcal{H}$ *be a set of subsets of* Q*. Let* (X, Y) *be a pair of random variables in* $Q \times \mathcal{Y}$*. We define:*

$$\mathcal{A}_{\mathcal{H},\delta}(X, Y) = \Big\{ y \in \mathcal{Y} : \ \exists H_y \in \mathcal{H}, \ ||\mathrm{P}_{X|Y=y} - \mathbb{I}_{H_y}||_\infty < \delta \Big\},$$

$$\mathcal{P}_{\mathcal{H},\delta}(X; Y) = \mathrm{P}_Y\big(\mathcal{A}_{\mathcal{H},\delta}(X, Y) \big).$$

If $\mathcal{P}_{\mathcal{H},\delta}(X; Y) > 1 - \delta$ for a small enough δ, then Y is "almost equivalent" to $\text{Proj}_{\mathcal{H}}(X)$. In the next lemma we will show that if $I(P^-)$ is close to $I(P)$, then the output of P is "almost equivalent" to the projection of the input onto a certain balanced partition $\mathcal{H}$.

Lemma 6. *Let* Q *and* $\mathcal{Y}$ *be two arbitrary sets with* $|Q| \geq 2$*. Let* (X, Y) *be a pair of random variables in* $Q \times \mathcal{Y}$ *such that* X *is uniform. Let* $\mathcal{H}$ *be a set of disjoint subsets of* Q *that have the same size. If* $\mathcal{P}_{\mathcal{H}, \frac{1}{|Q|^2}}(X; Y) > 1 - \frac{1}{|Q|^2}$*, then* $\mathcal{H}$ *is a balanced partition of* Q*.*

Proof: We only need to show that $\mathcal{H}$ covers Q. Suppose that there exists $x \in Q$ such that there is no H in $\mathcal{H}$ such that $x \in H$. Then for all $y \in \mathcal{A}_{\mathcal{H}, \frac{1}{|Q|^2}}(X, Y)$, $\mathrm{P}_{X|Y}(x|y) < \frac{1}{|Q|^2}$.

We have:

$$\begin{aligned}\mathrm{P}_X(x) = &\sum_{y\in\mathcal{A}_{\mathcal{H},\frac{1}{|Q|^2}}(X,Y)} \mathrm{P}_{X|Y}(x|y).\mathrm{P}_Y(y)\\ &+\sum_{y\in\mathcal{A}_{\mathcal{H},\frac{1}{|Q|^2}}(X,Y)^c} \mathrm{P}_{X|Y}(X|Y).\mathrm{P}_Y(Y)\\ &< \frac{1}{|Q|^2}\mathrm{P}_Y\big(\mathcal{A}_{\mathcal{H},\frac{1}{|Q|^2}}(X,Y)\big) + \mathrm{P}_Y\big(\mathcal{A}_{\mathcal{H},\frac{1}{|Q|^2}}(X,Y)^c\big)\\ &< \frac{1}{|Q|^2}+\frac{1}{|Q|^2} = \frac{2}{|Q|^2} \le \frac{1}{|Q|}.\end{aligned}$$

which is a contradiction since X is uniform in Q. Therefore, $\mathcal{H}$ covers Q and so it is a balanced partition of Q. ∎

Lemma 7. *Let Q and $\mathcal{Y}$ be two arbitrary sets with $|Q|\ge 2$, and let $\mathcal{H}$ and $\mathcal{H}'$ be two balanced partitions of Q. Let (X,Y) be a pair of random variables in $Q\times\mathcal{Y}$ such that X is uniform. If $\mathcal{P}_{\mathcal{H},\frac{1}{|Q|^2}}(X;Y) > 1-\frac{1}{2|Q|^2}$ and $\mathcal{P}_{\mathcal{H}',\frac{1}{|Q|^2}}(X;Y) > 1-\frac{1}{2|Q|^2}$, then $\mathcal{H}=\mathcal{H}'$.*

Proof: Define $\mathcal{H}''=\mathcal{H}\cap\mathcal{H}'$. Let $y\in\mathcal{A}_{\mathcal{H},\frac{1}{|Q|^2}}(X,Y)\cap\mathcal{A}_{\mathcal{H}',\frac{1}{|Q|^2}}(X,Y)$, choose $H\in\mathcal{H}$ and $H'\in\mathcal{H}'$ such that $||P_{X|Y=y}-\mathbb{I}_H||_\infty < \frac{1}{|Q|^2}$ and $||P_{X|Y=y}-\mathbb{I}_{H'}||_\infty<\frac{1}{|Q|^2}$, then

$$||\mathbb{I}_{H'}-\mathbb{I}_H||_\infty < \frac{2}{|Q|^2}\le\frac{1}{|Q|}$$

which implies that $H=H'$ and $y\in\mathcal{A}_{\mathcal{H}'',\frac{1}{|Q|^2}}(X,Y)$. Therefore,

$$\begin{aligned}\mathcal{P}_{\mathcal{H}'',\frac{1}{|Q|^2}}(X;Y) &\ge \mathrm{P}_Y\big(\mathcal{A}_{\mathcal{H},\frac{1}{|Q|^2}}(X,Y)\cap\mathcal{A}_{\mathcal{H}',\frac{1}{|Q|^2}}(X,Y)\big)\\ &> 1-\frac{1}{|Q|^2}.\end{aligned}$$

From *lemma 6* we conclude that $\mathcal{H}''$ is a balanced partition. Therefore, $\mathcal{H}=\mathcal{H}'=\mathcal{H}''$. ∎

Lemma 8. *Let $(Q,*)$ be a quasigroup with $|Q|\ge 2$, and let $\mathcal{Y}$ be an arbitrary set. For any $\delta>0$, there exists $\epsilon_1(\delta)>0$ depending only on Q such that for any two pairs of random variables (X_1,Y_1) and (X_2,Y_2) that are identically distributed in $Q\times\mathcal{Y}$ in such a way that X_1 and X_2 are uniform in Q, then $H(X_1*X_2|Y_1,Y_2)<H(X_1|Y_1)+\epsilon_1(\delta)$ implies the existence of a balanced partition $\mathcal{H}$ of Q such that $\mathcal{P}_{\mathcal{H},\delta}(X_1;Y_1)>1-\delta$ Moreover, $|H*H'|=|H|=|H'|$ for every $H,H'\in\mathcal{H}$.*

Proof: Choose $\delta>0$, and let $\delta'=\min\Big\{\frac{\delta}{|Q|^2},\frac{1}{|Q|^4}\Big\}$. Define:

- $p_{y_1}(x_1):=\mathrm{P}_{X_1|Y_1}(x_1|y_1)$ and $p_{y_1,x_2}(x):=p_{y_1}(x/^*x_2)$.
- $q_{y_2}(x_2):=\mathrm{P}_{X_2|Y_2}(x_2|y_2)$ and $q_{y_2,x_1}(x):=q_{y_2}(x_1\backslash_*x)$.

We have:

$$\begin{aligned}\mathrm{P}_{X_1*X_2|Y_1,Y_2}(x|y_1,y_2) &= \sum_{x_1\in Q} p_{y_1}(x_1)q_{y_2,x_1}(x) \quad (6)\\ &= \sum_{x_2\in Q} q_{y_2}(x_2)p_{y_1,x_2}(x). \quad (7)\end{aligned}$$

Due to the strict concavity of the entropy function, there exists $\epsilon'(\delta')>0$ such that:

- If $\exists x_1,x_1'\in Q$ such that $p_{y_1}(x_1)\ge\delta'$, $p_{y_1}(x_1')\ge\delta'$ and $||q_{y_2,x_1}-q_{y_2,x_1'}||_\infty\ge\delta'$ then

$$\begin{aligned}&H(X_1*X_2|Y_1=y_1,Y_2=y_2)\\ &\qquad\ge H(X_2|Y_2=y_2)+\epsilon'(\delta'), \quad (8)\end{aligned}$$

  (see (6)).
- If $\exists x_2,x_2'\in Q$ such that $q_{y_2}(x_2)\ge\delta'$, $q_{y_2}(x_2')\ge\delta'$ and $||p_{y_1,x_2}-p_{y_1,x_2'}||_\infty\ge\delta'$ then

$$\begin{aligned}&H(X_1*X_2|Y_1=y_1,Y_2=y_2)\\ &\qquad\ge H(X_1|Y_1=y_1)+\epsilon'(\delta'), \quad (9)\end{aligned}$$

  (see (7)).

Define:

$$\begin{aligned}\mathcal{C}_1=\Big\{&(y_1,y_2)\in\mathcal{Y}\times\mathcal{Y}:\ \forall x_1,x_1'\in Q,\\ &(p_{y_1}(x_1)\ge\delta',\ p_{y_1}(x_1')\ge\delta')\Rightarrow||q_{y_2,x_1}-q_{y_2,x_1'}||_\infty<\delta'\Big\},\end{aligned}$$

$$\begin{aligned}\mathcal{C}_2=\Big\{&(y_1,y_2)\in\mathcal{Y}\times\mathcal{Y}:\ \forall x_2,x_2'\in Q,\\ &(q_{y_2}(x_2)\ge\delta',\ q_{y_2}(x_2')\ge\delta')\Rightarrow||p_{y_1,x_2}-p_{y_1,x_2'}||_\infty<\delta'\Big\}.\end{aligned}$$

From (8) we have:

$$\begin{aligned}H(X_1*X_2|Y_1,Y_2)\ge\ &H(X_2|Y_2)+\epsilon'(\delta')\mathrm{P}_{Y_1,Y_2}(\mathcal{C}_1^c)\\ &= H(X_1|Y_1)+\epsilon'(\delta')\mathrm{P}_{Y_1,Y_2}(\mathcal{C}_1^c).\end{aligned}$$

Similarly, from (9) we have

$$H(X_1*X_2|Y_1,Y_2)\ge H(X_1|Y_1)+\epsilon'(\delta')\mathrm{P}_{Y_1,Y_2}(\mathcal{C}_2^c).$$

Let $\epsilon_1(\delta)=\epsilon'(\delta')\frac{\delta'^2}{2}$, and suppose that

$$H(X_1*X_2|Y_1,Y_2)<H(X_1|Y_1)+\epsilon_1(\delta),$$

then we must have $\mathrm{P}_{Y_1,Y_2}(\mathcal{C}_1^c)<\frac{\delta'^2}{2}$ and $\mathrm{P}_{Y_1,Y_2}(\mathcal{C}_2^c)<\frac{\delta'^2}{2}$, which imply that $\mathrm{P}_{Y_1,Y_2}(\mathcal{C})>1-\delta'^2$, where $\mathcal{C}=\mathcal{C}_1\cap\mathcal{C}_2$.

Now for each $a,a',x\in Q$, define:

- $\pi_{a,a'}(x):=(x*a)/^*a'$, and $\gamma_{a,a'}(x):=a'\backslash_*(a*x)$.

And for each $(y_1,y_2)\in\mathcal{Y}\times\mathcal{Y}$, define:

- $A_{y_1}:=\{x_1\in Q,p_{y_1}(x_1)\ge\delta'\}$.
- $B_{y_2}:=\{x_2\in Q,q_{y_2}(x_2)\ge\delta'\}$.
- $a_{y_1}:=\arg\max_{x_1}\ p_{y_1}(x_1)$. $b_{y_2}:=\arg\max_{x_2}\ q_{y_2}(x_2)$.
- $H_{y_1,y_2}=\Big\{x_1\in Q:\exists b_1,b_1',b_2,b_2',...,b_n,b_n'\in B_{y_2},\ x_1=(\pi_{b_n,b_n'}\circ...\circ\pi_{b_1,b_1'})(a_{y_1})\Big\}$.
- $K_{y_1,y_2}=\Big\{x_2\in Q:\exists a_1,a_1',a_2,a_2',...,a_n,a_n'\in A_{y_1},\ x_2=(\gamma_{a_n,a_n'}\circ...\circ\gamma_{a_1,a_1'})(b_{y_2})\Big\}$.

Suppose that $(y_1,y_2)\in\mathcal{C}$. Let $x_1\in H_{y_1,y_2}$, and let n be minimal such that there exists $b_1,b_1',b_2,b_2',...,b_n,b_n'\in B_{y_2}$ satisfying $x_1=(\pi_{b_n,b_n'}\circ...\circ\pi_{b_1,b_1'})(a_{y_1})$. Define $a_1:=a_{y_1}$,

and for $1 \leq i \leq n$ define $a_{i+1} = \pi_{b_i, b_i'}(a_i)$, so that $a_{n+1} = x_1$. We must have $a_i \neq a_j$ for $i \neq j$ since n was chosen to be minimal. Therefore, $n+1 \leq |Q|$.

For any $1 \leq i \leq n$, we have $a_{i+1} = (a_i * b_i)/^* b_i'$. Let $x = a_i * b_i$, then $a_{i+1} = x/^* b_i'$ and $a_i = x/^* b_i$. We have $(y_1, y_2) \in \mathcal{C}$, $q_{y_2}(b_i) \geq \delta'$ and $q_{y_2}(b_i') \geq \delta'$, so we must have $||p_{y_1, b_i'} - p_{y_1, b_i}||_\infty < \delta'$, and $|p_{y_1, b_i'}(x) - p_{y_1, b_i}(x)| < \delta'$, which implies that $|p_{y_1}(a_{i+1}) - p_{y_1}(a_i)| < \delta'$. Therefore:

$$\begin{aligned} &|p_{y_1}(x_1) - p_{y_1}(a_{y_1})| \\ &= |p_{y_1}(a_{n+1}) - p_{y_1}(a_1)| \leq \sum_{i=1}^{n} |p_{y_1}(a_{i+1}) - p_{y_1}(a_i)| \\ &< n\delta' \leq (|Q|-1)\delta' \leq \frac{|Q|-1}{|Q|^4} < \frac{|Q|-1}{|Q|^2}. \end{aligned} \tag{10}$$

Since $p_{y_1}(a_{y_1}) \geq \frac{1}{|Q|}$, we have $p_{y_1}(x_1) > \frac{1}{|Q|^2} > \delta'$ for every $x_1 \in H_{y_1, y_2}$. Therefore, $H_{y_1, y_2} \subset A_{y_1}$ $\forall (y_1, y_2) \in \mathcal{C}$. A similar argument yields $K_{y_1, y_2} \subset B_{y_2}$ $\forall (y_1, y_2) \in \mathcal{C}$.

Fix two elements $b, b' \in B_{y_2}$. We have $(x_1 * b)/^* b' \in H_{y_1, y_2}$ and so $x_1 * b \in H_{y_1, y_2} * b'$ for any $x_1 \in H_{y_1, y_2}$. Therefore, $H_{y_1, y_2} * b \subset H_{y_1, y_2} * b'$. But this is true for any two elements $b, b' \in B_{y_2}$, so $H_{y_1, y_2} * b = H_{y_1, y_2} * b'$ $\forall b, b' \in B_{y_2}$, and $|H_{y_1, y_2} * B_{y_2}| = |H_{y_1, y_2}|$. Similarly, we have $|A_{y_1} * K_{y_1, y_2}| = |K_{y_1, y_2}|$. If we also take into consideration the fact that $H_{y_1, y_2} \subset A_{y_1}$ and $K_{y_1, y_2} \subset B_{y_2}$ we conclude:

$$|B_{y_2}| \leq |H_{y_1, y_2} * B_{y_2}| = |H_{y_1, y_2}| \leq |A_{y_1}|,$$
$$|A_{y_1}| \leq |A_{y_1} * K_{y_1, y_2}| = |K_{y_1, y_2}| \leq |B_{y_2}|.$$

Therefore, $|A_{y_1}| = |H_{y_1, y_2}| = |B_{y_2}| = |K_{y_1, y_2}|$. We conclude that $H_{y_1, y_2} = A_{y_1}$ and $K_{y_1, y_2} = B_{y_2}$. Moreover, we have $|A_{y_1} * B_{y_2}| = |A_{y_1}| = |B_{y_2}|$.

Recall that $|p_{y_1}(x_1) - p_{y_1}(a_{y_1})| < (|Q|-1)\delta'$ for all $x_1 \in A_{y_1}$ (see (10)) and $p_{y_1}(x_1) < \delta' \leq (|Q|-1)\delta'$ for $x_1 \notin A_{y_1}$. It is easy to deduce that

$$||p_{y_1} - \mathbb{I}_{A_{y_1}}||_\infty < |Q|(|Q|-1)\delta' < |Q|^2\delta'.$$

Therefore, $||p_{y_1} - \mathbb{I}_{A_{y_1}}||_\infty < \delta$ and $||p_{y_1} - \mathbb{I}_{A_{y_1}}||_\infty < \frac{1}{|Q|^2}$. Similarly, $||q_{y_2} - \mathbb{I}_{B_{y_2}}||_\infty < \delta$ and $||q_{y_2} - \mathbb{I}_{B_{y_2}}||_\infty < \frac{1}{|Q|^2}$.

Now define $\mathcal{C}_{Y_1} = \Big\{ y_1 \in \mathcal{Y} :\ \mathrm{P}_{Y_2}\big((y_1, Y_2) \in \mathcal{C}\big) > 1 - \delta' \Big\}$, and for each $y_1 \in \mathcal{C}_{Y_1}$, define

$$\mathcal{K}_{y_1} = \Big\{ y_2 \in \mathcal{Y} :\ (y_1, y_2) \in \mathcal{C} \Big\}.$$

Then we have:

$$1 - \delta'^2 < \mathrm{P}_{Y_1, Y_2}(\mathcal{C}) \leq \Big(1 - \mathrm{P}_{Y_1}(\mathcal{C}_{Y_1})\Big)(1 - \delta') + \mathrm{P}_{Y_1}(\mathcal{C}_{Y_1}),$$

from which we conclude that $\mathrm{P}_{Y_1}(\mathcal{C}_{Y_1}) > 1 - \delta'$. And by definition, we also have $\mathrm{P}_{Y_2}(\mathcal{K}_{y_1}) > 1 - \delta'$ for all $y_1 \in \mathcal{C}_{Y_1}$. Define $\mathcal{H}_{y_1} = \{B_{y_2} :\ y_2 \in \mathcal{K}_{y_1}\}$.

Fix $y_1 \in \mathcal{C}_{Y_1}$. Since $|A_{y_1} * B| = |A_{y_1} * B'| = |A_{y_1}| = |B| = |B'|$ for every $B, B' \in \mathcal{H}_{y_1}$, we conclude that the elements of $\mathcal{H}_{y_1}$ are disjoint and have the same size (lemma 4). Now since $\mathrm{P}_{Y_2}(\mathcal{K}_{y_1}) > 1 - \frac{1}{|Q|^4}$ and since X_2 is uniform in Q, it is easy to see that $\mathcal{H}_{y_1}$ covers Q and so it is a balanced partition of Q for all $y_1 \in \mathcal{C}_{Y_1}$. Moreover, since $\mathrm{P}_{Y_2}(\mathcal{K}_{y_1}) > 1 - \frac{1}{|Q|^4}$, we can also conclude that all the balanced partitions $\mathcal{H}_{y_1}$ are the same. Let us denote this common balanced partition by $\mathcal{H}'$.

We have $|A * B| = |A| = |B|$ for all $A \in \mathcal{H}$ and all $B \in \mathcal{H}'$, where $\mathcal{H} = \{A_{y_1} :\ y_1 \in \mathcal{C}_{Y_1}\}$. By using a similar argument as in the previous paragraph, we can deduce that $\mathcal{H}$ is a balanced partition of Q. Moreover, since (X_1, Y_1) and (X_2, Y_2) are identically distributed, we can see that $\mathcal{H} = \mathcal{H}'$. We conclude the existence of a balanced partition $\mathcal{H}$ of Q satisfying $|A * B| = |A| = |B|$ for all $A, B \in \mathcal{H}$ and

$$\begin{aligned} &P_{Y_1}\left(\Big\{ y \in \mathcal{Y} :\ \exists H_y \in \mathcal{H},\ ||P_{X_1|Y_1=y} - \mathbb{I}_{H_y}||_\infty < \delta \Big\} \right) \\ &\quad \geq \mathrm{P}_{Y_1}(\mathcal{C}_{Y_1}) > 1 - \delta' > 1 - \delta. \end{aligned}$$

■

Lemma 9. *Let X_1 and X_2 be two independent random variables in Q such that there exists two sets $A, B \subset Q$ satisfying $||\mathrm{P}_{X_1} - \mathbb{I}_A||_\infty < \delta$, $||\mathrm{P}_{X_2} - \mathbb{I}_B||_\infty < \delta$ and $|A * B| = |A| = |B|$, then $||\mathrm{P}_{X_1 * X_2} - \mathbb{I}_{A*B}||_\infty < 2\delta + |Q|\delta^2$.*

Proof: The fact that $|A * B| = |A| = |B|$ implies that for every $x \in A * B$, we have $x/^* b \in A$ for every $b \in B$, and $x/^* b \in A^c$ for every $b \in B^c$.

For every $a \in Q$ define $\epsilon_{1,a} = \mathrm{P}_{X_1}(a) - \frac{1}{|A|}$ if $a \in A$, and $\epsilon_{1,a} = \mathrm{P}_{X_1}(a)$ if $a \notin A$. Similarly, for every $b \in Q$ define $\epsilon_{2,b} = \mathrm{P}_{X_2}(b) - \frac{1}{|A|}$ if $b \in B$, and $\epsilon_{2,b} = \mathrm{P}_{X_1}(b)$ if $b \notin B$. Let $x \in A * B$, we have:

$$\begin{aligned} &\mathrm{P}_{X_1 * X_2}(x) \\ &\quad = \sum_{b \in B} \mathrm{P}_{X_1}(x/^* b)\mathrm{P}_{X_2}(b) + \sum_{b \in B^c} \mathrm{P}_{X_1}(x/^* b)\mathrm{P}_{X_2}(b) \\ &\quad = \sum_{b \in B} \Big(\frac{1}{|A|} + \epsilon_{1, x/^* b}\Big)\Big(\frac{1}{|A|} + \epsilon_{2,b}\Big) + \sum_{b \in B^c} \epsilon_{1, x/^* b}\epsilon_{2,b} \\ &\quad = \frac{1}{|A|} + \frac{1}{|A|}\sum_{b \in B}\Big(\epsilon_{1, x/^* b} + \epsilon_{2,b}\Big) + \sum_{b \in Q} \epsilon_{1, x/^* b}\epsilon_{2,b}. \end{aligned}$$

Therefore,

$$\Big|\mathrm{P}_{X_1 * X_2}(x) - \frac{1}{|A|}\Big| < 2\delta + |Q|\delta^2.$$

Now let $x \notin A * B$, we have:

$$\begin{aligned} &\mathrm{P}_{X_1 * X_2}(x) \\ &\quad = \sum_{b \in B} \mathrm{P}_{X_1}(x/^* b)\mathrm{P}_{X_2}(b) + \sum_{\substack{b \notin B \\ x/^* b \in A}} \mathrm{P}_{X_1}(x/^* b)\mathrm{P}_{X_2}(b) \\ &\qquad\qquad + \sum_{\substack{b \notin B \\ x/^* b \notin A}} \mathrm{P}_{X_1}(x/^* b)\mathrm{P}_{X_2}(b) \\ &\quad = \sum_{b \in B} \epsilon_{1, x/^* b}\Big(\frac{1}{|A|} + \epsilon_{2,b}\Big) + \sum_{\substack{b \notin B \\ x/^* b \in A}} \Big(\frac{1}{|A|} + \epsilon_{1, x/^* b}\Big)\epsilon_{2,b} \\ &\qquad\qquad + \sum_{\substack{b \notin B \\ x/^* b \notin A}} \epsilon_{1, x/^* b}\epsilon_{2,b} \\ &\quad \leq 2\delta + |Q|\delta^2. \end{aligned}$$

■

Lemma 10. *Let $(Q,*)$ be a quasigroup with $|Q|\geq 2$, and let $\mathcal{Y}$ be an arbitrary set. For any $\delta>0$, there exists $\epsilon(\delta)>0$ depending only on Q and δ such that for any channel $P: Q\longrightarrow \mathcal{Y}$, $|I(P^-)-I(P)|<\epsilon(\delta)$ and $|I(P^{--})-I(P^-)|<\epsilon(\delta)$ implies the existence of a balanced partition $\mathcal{H}$ of Q such that $\mathcal{H}^{/^*}=\{H/^*H':\ H,H'\in\mathcal{H}\}$ is also a balanced partition of Q, $\mathcal{P}_{\mathcal{H},\delta}(X_1;Y_1)>1-\delta$, $\mathcal{P}_{\mathcal{H},\delta}(U_2;Y_1,Y_2,U_1)>1-\delta$ and $\mathcal{P}_{\mathcal{H}/^*,\delta}(U_1;Y_1,Y_2)>1-\delta$. Where U_1 and U_2 are two independent random variables uniformly distributed in Q, $X_1=U_1*U_2$, $X_2=U_2$, and Y_1 (resp. Y_2) is the output of the channel P when X_1 (resp. X_2) is the input.*

Proof: Let $\delta'=\min\{\delta,\delta'',\frac{1}{16|Q|^2}\}$, where $\delta''>0$ is a small enough number that will be specified later. Let $\epsilon(\delta)=\epsilon_1(\delta')$, where ϵ_1 is given by *lemma 8*. Let $P:Q\longrightarrow\mathcal{Y}$ be a channel as in the hypothesis. Then from *lemma 8* we conclude the existence of two balanced partitions $\mathcal{H}$ and $\mathcal{H}'$ such that $\mathcal{P}_{\mathcal{H},\delta'}(X_1;Y_1)>1-\delta'$ and $\mathcal{P}_{\mathcal{H}',\delta'}(U_1;Y_1,Y_2)>1-\delta'$. Moreover, we have $|H_1/^*H_2|=|H_1|=|H_2|$ for every $H_1,H_2\in\mathcal{H}$.

For each $H\in\mathcal{H}$, define:

$$\begin{aligned}A_H&=\Big\{y\in\mathcal{Y}:\ ||\mathrm{P}_{X_1|Y_1=y}-\mathbb{I}_H||_\infty<\delta'\Big\}\\&=\Big\{y\in\mathcal{Y}:\ ||\mathrm{P}_{X_2|Y_2=y}-\mathbb{I}_H||_\infty<\delta'\Big\},\end{aligned}$$

(note that (X_1,Y_1) and (X_2,Y_2) are identically distributed).

Let $x\in H$, we have:

$$\begin{aligned}\frac{1}{|Q|}&=\mathrm{P}_{X_1}(x)\\&=\sum_{y\in\mathcal{A}_{\mathcal{H},\delta'}(X_1;Y_1)\setminus A_H}\mathrm{P}_{X_1|Y_1}(x|y)\mathrm{P}_{Y_1}(y)\\&\qquad+\sum_{y\in A_h}\mathrm{P}_{X_1|Y_1}(x|y)\mathrm{P}_{Y_1}(y)\\&\qquad+\sum_{y\notin\mathcal{A}_{\mathcal{H},\delta'}(X_1;Y_1)}\mathrm{P}_{X_1|Y_1}(x|y)\mathrm{P}_{Y_1}(y)\\&\leq\delta'\mathcal{P}_{\mathcal{H},\delta'}(X_1;Y_1)+\Big(\frac{1}{|H|}+\delta'\Big)\mathrm{P}_{Y_1}(A_h)\\&\qquad+(1-\mathcal{P}_{\mathcal{H},\delta'}(X_1;Y_1))\\&<2\delta'+2\mathrm{P}_{Y_1}(A_H)\leq\frac{1}{8|Q|^2}+2\mathrm{P}_{Y_1}(A_H)\\&<\frac{1}{2|Q|}+2\mathrm{P}_{Y_1}(A_H).\end{aligned}$$

Therefore,

$$\mathrm{P}_{Y_2}(A_H)=\mathrm{P}_{Y_1}(A_H)>\frac{1}{4|Q|}.\tag{11}$$

Now for each $H_1,H_2\in\mathcal{H}$, define:

$$A'_{H_1,H_2}=\Big\{(y_1,y_2)\in\mathcal{Y}\times\mathcal{Y}:\ ||P_{U_1|Y_1=y_1,Y_2=y_2}-\mathbb{I}_{H_1/^*H_2}||_\infty<\frac{1}{2|Q|}\Big\}.$$

Let $(y_1,y_2)\in A_{H_1}\times A_{H_2}$, then $||\mathrm{P}_{X_1|Y_1=y_1}-\mathbb{I}_{H_1}||_\infty<\delta'$ and $||\mathrm{P}_{X_2|Y_2=y_2}-\mathbb{I}_{H_2}||_\infty<\delta'$. *Lemma 9* implies that

$$\begin{aligned}&||P_{U_1|Y_1=y_1,Y_2=y_2}-\mathbb{I}_{H_1/^*H_2}||_\infty\\&\quad=||P_{X_1/^*X_2|Y_1=y_1,Y_2=y_2}-\mathbb{I}_{H_1/^*H_2}||_\infty<2\delta'+|Q|\delta'^2\\&\quad\leq\frac{1}{8|Q|^2}+|Q|\frac{1}{16^2|Q|^4}<\frac{1}{2|Q|}.\end{aligned}$$

Therefore, $A_{H_1}\times A_{H_2}\subset A'_{H_1,H_2}$ and so $\mathrm{P}_{Y_1,Y_2}(A'_{H_1,H_2})\geq\mathrm{P}_{Y_1}(A_{H_1})\mathrm{P}_{Y_2}(A_{H_2})>\frac{1}{16|Q|^2}\geq\delta'$ (see (11)). We recall that $\mathrm{P}_{Y_1,Y_2}\big(\mathcal{A}_{\mathcal{H}',\delta'}(U_1;Y_1,Y_2)\big)=\mathcal{P}_{\mathcal{H}',\delta'}(U_1;Y_1,Y_2)>1-\delta'$, so $\mathcal{A}_{\mathcal{H}',\delta'}(U_1;Y_1,Y_2)\cap A'_{H_1,H_2}\neq\emptyset$.

Let $(y_1,y_2)\in\mathcal{A}_{\mathcal{H}',\delta'}(U_1;Y_1,Y_2)\cap A'_{H_1,H_2}$, then there exists $H'\in\mathcal{H}'$ such that $||\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2}-\mathbb{I}_{H'}||_\infty<\delta'<\frac{1}{2|Q|}$. Now since $(y_1,y_2)\in A'_{H_1,H_2}$, we have $||\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2}-\mathbb{I}_{H_1/^*H_2}||_\infty<\frac{1}{2|Q|}$, so $||\mathbb{I}_{H'}-\mathbb{I}_{H_1/^*H_2}||_\infty<\frac{1}{|Q|}$, we conclude that $H'=H_1/^*H_2$ and $H_1/^*H_2\in\mathcal{H}'$. But this is true for any $H_1,H_2\in\mathcal{H}$. Therefore, $\mathcal{H}^{/^*}\subset\mathcal{H}'$, which implies that $\mathcal{H}^{/^*}=\mathcal{H}'$ since both $\mathcal{H}'$ and $\mathcal{H}^{/^*}$ are partitions of Q whose all elements are non-empty. Thus,

$$\mathcal{P}_{\mathcal{H},\delta}(X_1;Y_1)\geq\mathcal{P}_{\mathcal{H},\delta'}(X_1;Y_1)>1-\delta'\geq1-\delta,$$

$$\mathcal{P}_{\mathcal{H}/^*,\delta}(U_1;Y_1,Y_2)\geq\mathcal{P}_{\mathcal{H}/^*,\delta'}(U_1;Y_1,Y_2)>1-\delta'\geq1-\delta.$$

It remains to prove that $\mathcal{P}_{\mathcal{H},\delta}(U_2;Y_1,Y_2,U_1)>1-\delta$. Define:

$$\mathcal{K}=\mathcal{A}_{\mathcal{H}/^*,\delta''}(U_1;Y_1,Y_2)\cap\Big(\mathcal{A}_{\mathcal{H},\delta''}(X_1;Y_1)\times\mathcal{A}_{\mathcal{H},\delta''}(X_2;Y_2)\Big).$$

We have:

$$\begin{aligned}\mathrm{P}_{Y_1}(\mathcal{A}_{\mathcal{H},\delta''}(X_1;Y_1))&=\mathrm{P}_{Y_2}(\mathcal{A}_{\mathcal{H},\delta''}(X_2;Y_2))\\&=\mathcal{P}_{\mathcal{H},\delta''}(X_1;Y_1)\geq\mathcal{P}_{\mathcal{H},\delta'}(X_1;Y_1)\\&>1-\delta'\geq1-\delta''.\end{aligned}$$

Thus, $\mathrm{P}_{Y_1,Y_2}(\mathcal{A}_{\mathcal{H},\delta''}(X_1;Y_1)\times\mathcal{A}_{\mathcal{H},\delta''}(X_2;Y_2))>1-2\delta''$. On the other hand, we have:

$$\begin{aligned}&\mathrm{P}_{Y_1,Y_2}(\mathcal{A}_{\mathcal{H}/^*,\delta''}(U_1;Y_1,Y_2))\\&\quad=\mathcal{P}_{\mathcal{H}/^*,\delta''}(U_1;Y_1,Y_2)=\mathcal{P}_{\mathcal{H}',\delta''}(U_1;Y_1,Y_2)\\&\quad\geq\mathcal{P}_{\mathcal{H}',\delta'}(U_1;Y_1,Y_2)>1-\delta'\geq1-\delta'',\end{aligned}$$

we conclude that $\mathrm{P}_{Y_1,Y_2}(\mathcal{K})>1-3\delta''$. Define:

$$\mathcal{B}=\Big\{(y_1,y_2,u_1)\in\mathcal{Y}\times\mathcal{Y}\times Q:\ (y_1,y_2)\in\mathcal{K},\text{and}\ \exists H\in\mathcal{H}^{/^*},\ ||\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2}-\mathbb{I}_H||_\infty<\delta''\text{ and }u_1\in H\Big\}.$$

If $(y_1,y_2)\in\mathcal{K}$, then $(y_1,y_2)\in\mathcal{A}_{\mathcal{H}/^*,\delta''}(U_1;Y_1,Y_2)$ and so there exists $H_{y_1,y_2}\in\mathcal{H}^{/^*}$ such that

$$||\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2}-\mathbb{I}_{H_{y_1,y_2}}||_\infty<\delta'',$$

which implies that $(y_1,y_2,u_1)\in\mathcal{B}$ for all $u_1\in H_{y_1,y_2}$. Now since $||\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2}-\mathbb{I}_{H_{y_1,y_2}}||_\infty<\delta''$, it is easy to see that $\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2}(H_{y_1,y_2})\geq1-|H_{y_1,y_2}|\delta''\geq1-|Q|\delta''$. Therefore,

$$\begin{aligned}\mathrm{P}_{Y_1,Y_2,U_1}(\mathcal{B})&>\mathrm{P}_{Y_1,Y_2}(\mathcal{K})(1-|Q|\delta'')\\&>(1-3\delta'')(1-|Q|\delta'')>1-(|Q|+3)\delta''.\end{aligned}$$

Therefore, if $\delta'' \leq \frac{\delta}{|Q|+3}$, then $\mathrm{P}_{Y_1,Y_2,U_1}(\mathcal{B}) > 1-\delta$.

Now let $(y_1, y_2, u_1) \in \mathcal{B}$. There exists $H_1, H_2 \in \mathcal{H}$ and $H \in \mathcal{H}^{/^*}$ such that:

- $u_1 \in H$,
- $||\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2} - \mathbb{I}_H||_\infty < \delta''$,
- $||\mathrm{P}_{X_1|Y_1=y_1} - \mathbb{I}_{H_1}||_\infty < \delta''$,
- $||\mathrm{P}_{X_2|Y_2=y_2} - \mathbb{I}_{H_2}||_\infty < \delta''$.

Since $U_1 = X_1 /^* X_2$, *lemma 9* implies that $||\mathrm{P}_{U_1|Y_1=y_1,Y_2=y_2} - \mathbb{I}_{H_1/^*H_2}||_\infty < 2\delta'' + |Q|\delta''^2$, and $||\mathbb{I}_H - \mathbb{I}_{H_1/^*H_2}||_\infty < 3\delta'' + |Q|\delta''^2$. Therefore, if $\delta'' \leq \frac{1}{4|Q|}$, then $||\mathbb{I}_H - \mathbb{I}_{H_1/^*H_2}||_\infty < \frac{1}{|Q|}$ and $H = H_1 /^* H_2$. Now we have:

- $u_1 \in H$ implies $\left|\mathrm{P}_{U_1|Y_1,Y_2}(u_1|y_1,y_2) - \frac{1}{|H|}\right| < \delta''$, i.e., $\frac{1}{|H|} - \delta'' < \mathrm{P}_{U_1|Y_1,Y_2}(u_1|y_1,y_2) < \frac{1}{|H|} + \delta''$.
- If $u_2 \in H_2$, then $u_1 * u_2 \in H_1$ which implies that $\left|\mathrm{P}_{X_1|Y_1}(u_1 * u_2|y_1) - \frac{1}{|H|}\right| < \delta''$ and $\left|\mathrm{P}_{X_2|Y_2}(u_2|y_2) - \frac{1}{|H|}\right| < \delta''$.
- If $u_2 \notin H_2$, then $u_1 * u_2 \notin H_1$, so $\mathrm{P}_{X_1|Y_1}(u_1 * u_2|y_1) < \delta''$ and $\mathrm{P}_{X_2|Y_2}(u_2|y_2) < \delta''$.

By noticing that

$$\begin{aligned}\mathrm{P}_{U_2|Y_1,Y_2,U_1}(u_2|y_1,y_2,u_1) &= \frac{\mathrm{P}_{U_2,U_1|Y_1,Y_2}(u_2,u_1|y_1,y_2)}{\mathrm{P}_{U_1|Y_1,Y_2}(u_1|y_1,y_2)} \\ &= \frac{\mathrm{P}_{X_1|Y_1}(u_1 * u_2|y_1)\mathrm{P}_{X_2|Y_2}(u_2|y_1)}{\mathrm{P}_{U_1|Y_1,Y_2}(u_1|y_1,y_2)},\end{aligned}$$

we conclude that:

- If $u_2 \in H_2$, we have:

$$\frac{\left(\frac{1}{|H|} - \delta''\right)^2}{\frac{1}{|H|} + \delta''} < \mathrm{P}_{U_2|Y_1,Y_2,U_1}(u_2|y_1,y_2,u_1) < \frac{\left(\frac{1}{|H|} + \delta''\right)^2}{\frac{1}{|H|} - \delta''}.$$

- If $u_2 \notin H_2$, we have:

$$\mathrm{P}_{U_2|Y_1,Y_2,U_1}(u_2|y_1,y_2,u_1) < \frac{\delta''^2}{\frac{1}{|H|} - \delta''}.$$

Now since $\lim_{\delta''\to 0} \frac{\left(\frac{1}{|H|} - \delta''\right)^2}{\frac{1}{|H|} + \delta''} = \lim_{\delta''\to 0} \frac{\left(\frac{1}{|H|} + \delta''\right)^2}{\frac{1}{|H|} - \delta''} = \frac{1}{|H_2|}$ (as $|H| = |H_2|$) and since $\lim_{\delta''\to 0} \frac{\delta''^2}{\frac{1}{|H|} - \delta''} = 0$, there exists $\beta(\delta) > 0$ such that if $\delta'' \leq \beta(\delta)$ we get

$$||\mathrm{P}_{U_2|Y_1=y_1,Y_2=y_2,U_1=u_1} - \mathbb{I}_{H_2}||_\infty < \delta.$$

By setting $\delta'' = \min\left\{\frac{\delta}{|Q|+3}, \frac{1}{4|Q|}, \beta(\delta)\right\}$, we get $(y_1,y_2,u_1) \in \mathcal{A}_{\mathcal{H},\delta}(U_2; Y_1, Y_2, U_1)$ for every $(y_1,y_2,u_1) \in \mathcal{B}$, i.e., $\mathcal{B} \subset \mathcal{A}_{\mathcal{H},\delta}(U_2; Y_1, Y_2, U_1)$ and

$$\mathcal{P}_{\mathcal{H},\delta}(U_2; Y_1, Y_2, U_1) \geq \mathrm{P}_{Y_1,Y_2,U_1}(\mathcal{B}) > 1 - \delta.$$

■

Now we are ready to prove *theorem* 2. In fact, we will prove a stronger theorem:

Theorem 3. *Let $(Q,*)$ be a quasigroup and let $P : Q \longrightarrow \mathcal{Y}$ be an arbitrary channel. Then for any $\delta > 0$, we have:*

$$\lim_{n\to\infty} \frac{1}{2^n}\left|\left\{s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (Q,/^*), \left|I(P^s[\mathcal{H}']) - \log\frac{|\mathcal{H}_s|.||\mathcal{H}_s \wedge \mathcal{H}'||}{||\mathcal{H}'||}\right| < \delta \text{ for all stable partitions } \mathcal{H}' \text{ of } (Q,/^*)\right\}\right| = 1$$

Proof: Due to the continuity of the entropy function, and because of *lemma 3*, there exists $\gamma(\delta) > 0$ depending only on Q such that if (X,Y) is a pair of random variables in $Q \times \mathcal{Y}$ where X is uniform, and if there exists a stable partition of $\mathcal{H}$ such that $\mathcal{P}_{\mathcal{H},\gamma(\delta)}(X;Y) > 1 - \gamma(\delta)$, then $\left|I(X;Y) - \log|\mathcal{H}|\right| < \delta$ and $\left|I\big(\mathrm{Proj}_{\mathcal{H}'}(X); Y\big) - \log\frac{|\mathcal{H}|.||\mathcal{H}\wedge\mathcal{H}'||}{||\mathcal{H}'||}\right| < \delta$ for all stable partitions $\mathcal{H}'$ of $(Q,/^*)$.

Let P^n be as in *definition 13*. From *remark 2* we have:

$$\mathbb{E}\big(I(P_{n+1})|P_n\big) = \frac{1}{2}I(P_n^-) + \frac{1}{2}I(P_n^+) = I(P_n)$$

This implies that the process $\{I(P_n)\}_n$ is a martingale, and so it converges almost surely.

Let m be the number of different balanced partitions of Q, choose $l > m$ and let $0 \leq i \leq l+1$. Almost surely, $|I(P_{n-l+i+1}) - I(P_{n-l+i})|$ converges to zero. Therefore, we have:

$$\lim_{n\to\infty} \frac{1}{2^{n-l+i}}|A_{n,l,i}| = 1$$

where

$$A_{n,l,i} := \Big\{(s_1,s_2) \in \{-,+\}^{n-l} \times \{-,+\}^i : |I(P^{(s_1,s_2,-)}) - I(P^{(s_1,s_2)})| < \epsilon(\delta')\Big\},$$

and $\epsilon(\delta')$ is given by *lemma 10*. Now for each $s_2 \in \{-,+\}^i$, define:

$$A_{n,l,s_2} := \Big\{s_1 \in \{-,+\}^{n-l} : |I(P^{(s_1,s_2,-)}) - I(P^{(s_1,s_2)})| < \epsilon(\delta')\Big\}.$$

It is easy to see that $|A_{n,l,i}| = \sum_{s_2\in\{-,+\}^i} |A_{n,l,s_2}|$. Therefore,

$$\frac{1}{2^i}\sum_{s_2\in\{-,+\}^i}\Big(\lim_{n\to\infty}\frac{1}{2^{n-l}}|A_{n,l,s_2}|\Big) = \lim_{n\to\infty}\frac{1}{2^{n-l+i}}|A_{n,l,i}| = 1,$$

i.e.,

$$\sum_{s_2\in\{-,+\}^i}\Big(\lim_{n\to\infty}\frac{1}{2^{n-l}}|A_{n,l,s_2}|\Big) = 2^i. \tag{12}$$

On the other hand, it is obvious that $|A_{n,l,s_2}| \leq 2^{n-l}$, and so $\lim_{n\to\infty} \frac{1}{2^{n-l}}|A_{n,l,s_2}| \leq 1$ for all $s_2 \in \{-,+\}^i$. We can now use (12) to conclude that $\lim_{n\to\infty} \frac{1}{2^{n-l}}|A_{n,l,s_2}| \leq 1$ for all $s_2 \in \{-,+\}^i$. Therefore, we must have $\lim_{n\to\infty} \frac{1}{2^{n-l}}|A_{n,l}| = 1$, where

$$\begin{aligned} A_{n,l} &:= \bigcap_{\substack{0\leq i\leq l+1 \\ s_2\in\{-,+\}^i}} A_{n,l,s_2} \\ &= \Big\{ s_1 \in \{-,+\}^{n-l} : \ |I(P^{(s_1,s_2,-)}) - I(P^{(s_1,s_2)})| < \epsilon(\delta'), \\ &\qquad\qquad \forall s_2 \in \{-,+\}^i, \ \forall 0 \leq i \leq l+1 \Big\}. \end{aligned}$$

Now define:

$$\begin{aligned} C_l := \Big\{ & s_2 \in \{-,+\}^l : \\ & s_2 \text{ contains the sign } - \text{ at least } m \text{ times} \Big\}, \end{aligned}$$

$$\begin{aligned} B_{n,l} :&= A_{n,l} \times C_l \\ &= \Big\{ s = (s_1, s_2) \in \{-,+\}^{n-l} \times \{-,+\}^l : \\ &\qquad\qquad s_1 \in A_{n,l}, \ s_2 \in C_l \Big\}, \end{aligned}$$

$$\begin{aligned} D_n := \Big\{ & s \in \{-,+\}^n : \ \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ & \Big| I(P^s[\mathcal{H}']) - \log \frac{|\mathcal{H}_s|.||\mathcal{H}_s \wedge \mathcal{H}'||}{||\mathcal{H}'||} \Big| < \delta \\ & \qquad \text{for all stable partitions } \mathcal{H}' \text{ of } (Q, /^*) \Big\}. \end{aligned} \tag{13}$$

Now let $s_1 \in A_{n,l}$, let $n - l \leq j \leq n$, let $s = (s_1, s_2) \in \{-,+\}^j$ for some $s_2 \in \{-,+\}^{j-n+l}$, let X_s be the input to the channel P^s and Y_s be the output of it. Since $j - n + l \leq l$, both s_2 and $(s_2, -)$ have lengths of at most $l+1$. Therefore, we have $|I(P^{(s_1,s_2,-)}) - I(P^{(s_1,s_2)})| < \epsilon(\delta')$ and $|I(P^{(s_1,s_2,-,-)}) - I(P^{(s_1,s_2,-)})| < \epsilon(\delta')$. Lemma 10 implies the existence of a balanced partitions $\mathcal{H}_s$ such that $\mathcal{P}_{\mathcal{H}_s,\delta'}(X_s; Y_s) > 1 - \delta'$, $\mathcal{P}_{\mathcal{H}_s^{/^*},\delta'}(X_{(s,-)}; Y_{(s,-)}) > 1 - \delta'$ and $\mathcal{P}_{\mathcal{H}_s,\delta'}(X_{(s,+)}; Y_{(s,+)}) > 1 - \delta'$ for all $s \in \{-,+\}^j$ $(n - l \leq j \leq n)$ having s_1 as a prefix. Since $\delta' < \frac{1}{2|Q|^2}$, lemma 7 implies that $\mathcal{H}_{(s,-)} = \mathcal{H}_s^{/^*}$ and $\mathcal{H}_{(s,+)} = \mathcal{H}_s$ for all $s \in \{-,+\}^j$ $(n - l \leq j < n)$ having s_1 as a prefix.

Let $s_2 \in C_l$, and let l' be the number of $-$ signs in s_2 (we have $m \leq l' \leq l$), then there exist $l' + 1$ balanced partitions $\mathcal{H}_i$ $(0 \leq i \leq l')$ such that $\mathcal{H}_0 = \mathcal{H}_{s_1}$, $\mathcal{H}_{l'} = \mathcal{H}_{(s_1,s_2)}$, and $\mathcal{H}_{i+1} = \mathcal{H}_i^{/^*}$ for each $0 \leq i \leq l' - 1$. Since m is the number of different balanced partitions of Q, there exist two indices i and j such that $i < j \leq l'$ and $\mathcal{H}_i = \mathcal{H}_j$. We conclude that $\mathcal{H}_{l'} = \mathcal{H}_{(s_1,s_2)}$ is a stable partition of $(Q, /^*)$. Moreover, since $\delta' \leq \gamma(\delta)$, (s_1, s_2) belongs to D_n. Therefore, $B_{n,l} \subset D_n$ for any $l \geq m$. Thus:

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{2^n}|D_n| &\geq \lim_{n\to\infty} \frac{1}{2^n}|B_{n,l}| \\ &= \lim_{n\to\infty} \Big(\frac{1}{2^{n-l}}|A_{n,l}|\Big)\Big(\frac{1}{2^l}|C_l|\Big) = \frac{1}{2^l}|C_l|. \end{aligned}$$

But this is true for any $l \geq m$, we conclude:

$$\liminf_{n\to\infty} \frac{1}{2^n}|D_n| \geq \lim_{l\to\infty} \frac{1}{2^l}|C_l| = 1,$$

which implies that

$$\lim_{n\to\infty} \frac{1}{2^n}|D_n| = 1.$$

■

IV. Rate of polarization

In this section, we are interested in the rate of polarization of P_n to deterministic projection channels.

Definition 16. *The* Battacharyya parameter *of an ordinary channel P with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is defined as:*

$$Z(P) = \frac{1}{|\mathcal{X}|(|\mathcal{X}| - 1)} \sum_{\substack{(x,x')\in\mathcal{X}\times\mathcal{X} \\ x\neq x'}} \sum_{y\in\mathcal{Y}} \sqrt{P(y|x)P(y|x')}$$

if $|\mathcal{X}| > 1$. And by convention, we take $Z(P) = 0$ if $|\mathcal{X}| = 1$.

It's known that $\mathrm{P}_e(P) \leq |\mathcal{X}|Z(P)$ (see [3]), where $\mathrm{P}_e(P)$ is the probability of error of the maximum likelihood decoder of P.

Definition 17. *Let $(Q, *)$ be a quasigroup with $|Q| \geq 2$, and $\mathcal{Y}$ be an arbitrary set. Let $P : Q \longrightarrow \mathcal{Y}$ be an arbitrary channel, and $\mathcal{H}$ be a stable partition of $(Q, /^*)$. We define the channels $P[\mathcal{H}]^- : \mathcal{H}^{/^*} \longrightarrow \mathcal{Y} \times \mathcal{Y}$ and $P[\mathcal{H}]^+ : \mathcal{H} \longrightarrow \mathcal{Y} \times \mathcal{Y} \times \mathcal{H}^{/^*}$ by:*

$$P[\mathcal{H}]^+(y_1, y_2, H_1|H_2) = \frac{1}{|\mathcal{H}|} P[\mathcal{H}](y_1|H_1 * H_2) P[\mathcal{H}](y_2|H_2),$$

$$P[\mathcal{H}]^-(y_1, y_2|H_1) = \frac{1}{|\mathcal{H}|} \sum_{H_2\in\mathcal{H}} P[\mathcal{H}](y_1|H_1 * H_2) P[\mathcal{H}](y_2|H_2).$$

Lemma 11. *$P[\mathcal{H}]^+$ is degraded with respect to $P^+[\mathcal{H}]$, and $P[\mathcal{H}]^-$ is equivalent to $P^-[\mathcal{H}^{/^*}]$.*

Proof: Let $(H_1, H_2, y_1, y_2) \in \mathcal{H}^{/^*} \times \mathcal{H} \times \mathcal{Y} \times \mathcal{Y}$, we have:

$$
\begin{aligned}
&P[\mathcal{H}]^+(y_1, y_2, H_1|H_2) \\
&= \frac{1}{|\mathcal{H}|} P[\mathcal{H}](y_1|H_1 * H_2) P[\mathcal{H}](y_2|H_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}}(x_1) = H_1 * H_2}} P(y_1|x_1) \sum_{\substack{x_2 \in Q \\ \text{Proj}_{\mathcal{H}}(x_2) = H_2}} P(y_2|x_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{\substack{x_2 \in Q \\ \text{Proj}_{\mathcal{H}}(x_2) = H_2}} P(y_1|x_1 * x_2) P(y_2|x_2) \\
&= \frac{1}{||\mathcal{H}||} \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{\substack{x_2 \in Q \\ \text{Proj}_{\mathcal{H}}(x_2) = H_2}} P^+(y_1, y_2, x_1|x_2) \\
&= \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} P^+[\mathcal{H}](y_1, y_2, x_1|H_2).
\end{aligned}
$$

Therefore, $P[\mathcal{H}]^+$ is degraded with respect to $P^+[\mathcal{H}]$. Now let $(H_1, y_1, y_2) \in \mathcal{H}^{/^*} \times \mathcal{Y} \times \mathcal{Y}$, we have:

$$
\begin{aligned}
&P[\mathcal{H}]^-(y_1, y_2|H_1) \\
&= \frac{1}{|\mathcal{H}|} \sum_{H_2 \in \mathcal{H}} P[\mathcal{H}](y_1|H_1 * H_2) P[\mathcal{H}](y_2|H_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{H_2 \in \mathcal{H}} \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}}(x_1) = H_1 * H_2}} P(y_1|x_1) \sum_{\substack{x_2 \in Q \\ \text{Proj}_{\mathcal{H}}(x_2) = H_2}} P(y_2|x_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{H_2 \in \mathcal{H}} \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{\substack{x_2 \in Q \\ \text{Proj}_{\mathcal{H}}(x_2) = H_2}} P(y_1|x_1 * x_2) P(y_2|x_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{x_2 \in Q} P(y_1|x_1 * x_2) P(y_2|x_2) \\
&= \frac{1}{||\mathcal{H}||} \sum_{\substack{x_1 \in Q \\ \text{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} P^-(y_1, y_2|x_1) = P^-[\mathcal{H}^{/^*}](y_1, y_2|H_1).
\end{aligned}
$$

Therefore, $P[\mathcal{H}]^-$ is equivalent to $P^-[\mathcal{H}^{/^*}]$. ∎

Definition 18. *Let $\mathcal{H}$ be a stable partition of $(Q, /^*)$, we define the stable partitions $\mathcal{H}^-$ and $\mathcal{H}^+$, by $\mathcal{H}^{/^*}$ and $\mathcal{H}$ respectively.*

Lemma 12. *Let B_n and P_n be defined as in* definition 13. *For each stable partition $\mathcal{H}$ of $(Q, /^*)$, we define the stable partition-valued process $\mathcal{H}_n$ by:*

$$
\begin{aligned}
&\mathcal{H}_0 := \mathcal{H}, \\
&\mathcal{H}_n := \mathcal{H}_{n-1}^{B_n} \ \forall n \geq 1.
\end{aligned}
$$

Then $I(P_n[\mathcal{H}_n])$ converges almost surely to a number in $\mathcal{L}_{\mathcal{H}} := \big\{ \log d : \ d \text{ divides } |\mathcal{H}| \big\}$.

Proof: Since $P_n[\mathcal{H}_n]^-$ is equivalent to $P_n^-[\mathcal{H}_n^{/^*}]$ and $P_n[\mathcal{H}_n]^+$ is degraded with respect to $P_n^+[\mathcal{H}_n]$ (*lemma 11*), we have:

$$
\begin{aligned}
\mathbb{E}\Big(I(P_{n+1}[\mathcal{H}_{n+1}]) \Big| P_n \Big) &= \frac{1}{2} I(P_n^-[\mathcal{H}_n^{/^*}]) + \frac{1}{2} I(P_n^+[\mathcal{H}_n]) \\
&\geq \frac{1}{2} I(P_n[\mathcal{H}_n]^-) + \frac{1}{2} I(P_n[\mathcal{H}_n]^+) = I(P_n[\mathcal{H}_n]).
\end{aligned}
$$

This implies that the process $I(P_n[\mathcal{H}_n])$ is a sub-martingale and therefore it converges almost surely. Let $\delta > 0$, and define $D_{l,\delta}$ as in (13), we have shown that $\lim\limits_{n \longrightarrow \infty} \frac{1}{2^n} |D_{n,\delta}| = 1$. It is easy to see that almost surely, for every $\delta > 0$ and for every $n_0 > 0$ there exists $n > n_0$ such that $(B_1, ..., B_n) \in D_{l,\delta}$.

Let B_n be a realization in which $I(P_n[\mathcal{H}_n])$ converges to a limit x, and in which for every $\delta > 0$ and for every $n_0 > 0$ there exists $n > n_0$ such that $(B_1, ..., B_n) \in D_{n,\delta}$. Let $\delta > 0$ and let $n_0 > 0$ be chosen such that $|I(P_n[\mathcal{H}_n]) - x| < \delta$ for every $n > n_0$. Choose $n > n_0$ such that $(B_1, ..., B_n) \in D_{n,\delta}$, this means that there exists a stable partition $\mathcal{H}'$ of $(Q, /^*)$ such that

$$
\Big| I(P_n[\mathcal{H}_n]) - \log \frac{|\mathcal{H}'|.||\mathcal{H}' \wedge \mathcal{H}_n||}{||\mathcal{H}_n||} \Big| < \delta.
$$

Therefore, $\Big| x - \log \frac{|\mathcal{H}_n|.||\mathcal{H}' \wedge \mathcal{H}_n||}{||\mathcal{H}'||} \Big| < 2\delta$, which implies that $\Big| x - \log \frac{|\mathcal{H}'|.||\mathcal{H}' \wedge \mathcal{H}_n||}{||\mathcal{H}_n||} \Big|$ since $|Q| = |\mathcal{H}'|.||\mathcal{H}'|| = |\mathcal{H}_n|.||\mathcal{H}_n||$.

By noticing that $\frac{|\mathcal{H}_n|.||\mathcal{H}' \wedge \mathcal{H}_n||}{||\mathcal{H}'||}$ divides $|\mathcal{H}_n| = |\mathcal{H}|$, we conclude that $d(x, \mathcal{L}_{\mathcal{H}}) < 2\delta$ for every $\delta > 0$. Therefore, $x \in \mathcal{L}_{\mathcal{H}}$. ∎

Lemma 13. *Let $P : Q \to \mathcal{Y}$ be an ordinary channel where Q is a quasigroup with $|Q| \geq 2$. For any stable partition $\mathcal{H}$ of $(Q, /^*)$, we have:*

$$
\lim_{n \to \infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-, +\}^n : \ \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\
I(P^s[\mathcal{H}]) > \log |\mathcal{H}| - \epsilon, Z(P^s[\mathcal{H}]) \geq 2^{-2^{n\beta}} \Big\} \Big| = 0,
$$

for any $0 < \epsilon < \log 2$ and any $0 < \beta < \frac{1}{2}$.

Proof: Let $0 < \epsilon < \log 2$ and $0 < \beta < \frac{1}{2}$, and let $\mathcal{H}$ be a stable partition of $(Q, /^*)$. $I(P_n[\mathcal{H}_n])$ converges almost surely to an element in $\mathcal{L}_{\mathcal{H}}$. Due to the relations between the quantities $I(P)$ and $Z(P)$ (see *proposition 3.3* of [11]) we can see that $Z(P_n[\mathcal{H}_n])$ converges to 0 if and only if $I(P_n[\mathcal{H}_n])$ converges to $\log |\mathcal{H}|$, and there is a number $z_0 > 0$ such that $\liminf Z(P_n[H]) > z_0$ whenever $I(P_n[H])$ converges to a number in $\mathcal{L}_{\mathcal{H}}$ other than $\log |\mathcal{H}|$. Therefore, we can say that almost surely, we have:

$$
\lim Z(P_n[\mathcal{H}_n]) = 0 \ \text{ or } \ \liminf Z(P_n[H]) > z_0
$$

$Z(P_n^+[\mathcal{H}_n^+]) \leq Z(P_n[\mathcal{H}_n]^+)$ since $P_n[\mathcal{H}_n]^+$ is degraded with respect to $P_n^+[\mathcal{H}_n^+]$, and $Z(P_n^-[\mathcal{H}_n^-]) = Z(P_n[\mathcal{H}_n]^-)$

since $P_n[\mathcal{H}_n]^-$ and $P_n^-[\mathcal{H}_n^-]$ are equivalent (see *lemma* 11). From *lemma* 3.5 of [11] we have:

- $Z(P_n[\mathcal{H}_n]^-) \leq \big(|\mathcal{H}|^2 - |\mathcal{H}| + 1\big) Z(P_n[\mathcal{H}_n])$.
- $Z(P_n[\mathcal{H}_n]^+) \leq \big(|\mathcal{H}| - 1\big) Z(P_n[\mathcal{H}_n])^2$.

Therefore, we have $Z(P_n^-[\mathcal{H}_n]) \leq K.Z(P_n[\mathcal{H}_n])$ and $Z(P_n^+[\mathcal{H}_n]) \leq K.Z(P_n[\mathcal{H}_n])^2$, where K is equal to $\big(|\mathcal{H}|^2 - |\mathcal{H}| + 1\big)$. By applying exactly the same techniques that were used to prove *theorem 3.5* of [11] we get:

$$\lim_{n\to\infty} \Pr\Big(I(P_n[\mathcal{H}_n]) > \log|\mathcal{H}| - \epsilon, Z(P_n[\mathcal{H}_n]) \geq 2^{-2^{n\beta}}\Big) = 0$$

But this is true for all stable partitions $\mathcal{H}$. Therefore,

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ I(P^s[\mathcal{H}^s]) > \log|\mathcal{H}| - \epsilon, Z(P^s[\mathcal{H}^s]) \geq 2^{-2^{n\beta}} \Big\} \Big| = 0.$$

By noticing that for each $s \in \{-,+\}^n$, there exists a stable partition $\mathcal{H}_s$ such that $\mathcal{H} = \mathcal{H}_s^s$, we conclude:

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ I(P^s[\mathcal{H}]) > \log|\mathcal{H}| - \epsilon, Z(P^s[\mathcal{H}]) \geq 2^{-2^{n\beta}} \Big\} \Big| = 0.$$

■

Theorem 4. *The convergence of P_n to projection channels is almost surely fast:*

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ \big|I(P^s) - \log|\mathcal{H}|\big| < \epsilon, \big|I(P^s[\mathcal{H}]) - \log|\mathcal{H}|\big| < \epsilon, \\ Z(P^s[\mathcal{H}]) < 2^{-2^{\beta n}} \Big\} \Big| = 1,$$

for any $0 < \epsilon < \log 2$, and any $0 < \beta < \frac{1}{2}$.

Proof: Let $0 < \epsilon < \log 2$, and $0 < \beta < \frac{1}{2}$. Define:

$$E_0 = \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ I(P^s[\mathcal{H}]) > \log|\mathcal{H}| - \epsilon,\ Z(P^s[\mathcal{H}]) \geq 2^{-2^{\beta n}} \Big\},$$

$$E_1 = \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ \big|I(P^s) - \log|\mathcal{H}|\big| < \epsilon, \big|I(P^s[\mathcal{H}]) - \log|\mathcal{H}|\big| < \epsilon \Big\},$$

$$E_2 = \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ \big|I(P^s) - \log|\mathcal{H}|\big| < \epsilon, \big|I(P^s[\mathcal{H}]) - \log|\mathcal{H}|\big| < \epsilon, \\ Z(P^s[\mathcal{H}]) < 2^{-2^{\beta n}} \Big\}.$$

It is easy to see that $E_1 \setminus E_0 \subset E_2$ and $|E_2| \geq |E_1| - |E_0|$. By *theorem* 2 and *lemma* 13 we get:

$$1 \geq \lim_{n\to\infty} \frac{1}{2^n}|E_2| \geq \lim_{n\to\infty} \frac{1}{2^n}\big(|E_1| - |E_0|\big) = 1 - 0 = 1.$$

■

V. Polar codes construction

Choose $0 < \epsilon < \log 2$ and $0 < \beta < \beta' < \frac{1}{2}$, let n be an integer such that

$$|Q| 2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}} \text{ and } \frac{1}{2^n}|E_n| > 1 - \frac{\epsilon}{2\log|Q|},$$

where

$$E_n = \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ \big|I(P^s) - \log|\mathcal{H}|\big| < \frac{\epsilon}{2}, \big|I(P^s[\mathcal{H}]) - \log|\mathcal{H}|\big| < \frac{\epsilon}{2}, \\ Z(P^s[\mathcal{H}]) < 2^{-2^{\beta' n}} \Big\}.$$

Such an integer exists due to *theorem 4*. A polar code is constructed as follows: If $s \notin E_n$, let U_s be a frozen symbol, i.e., we suppose that the receiver knows U_s. On the other hand, if $s \in E_n$, there exists a stable partition $\mathcal{H}_s$ of G, such that $\big|I(P^s) - \log|\mathcal{H}_s|\big| < \frac{\epsilon}{2}$, $\big|I(P^s[\mathcal{H}_s]) - \log|\mathcal{H}_s|\big| < \frac{\epsilon}{2}$, and $Z(P^s[\mathcal{H}_s]) < 2^{-2^{\beta' n}}$. Let $f_s : \mathcal{H}_s \longrightarrow G$ be a frozen mapping (in the sense that the receiver knows f_s) such that $f_s(H) \in H$ for all $H \in \mathcal{H}_s$, we call such mapping *a section mapping*. We choose U'_s uniformly in $\mathcal{H}_s$ and we let $U_s = f_s(U'_s)$. Note that if the receiver can determine $\text{Proj}_{\mathcal{H}_s}(U_s) = U'_s$ accurately, then he can also determine U_s since he knows f_s.

Since we are free to choose any value for the frozen symbols and for the section mappings, we will analyse the performance of the polar code averaged on all the possible choices of the frozen symbols and for the section mappings. Therefore, U_s are independent random variables, uniformly distributed in Q. If $s \notin E_n$, the receiver knows U_s and there is nothing to decode, and if $s \in E_n$, the receiver has to determine $\text{Proj}_{\mathcal{H}_s}(U_s)$ in order to successfully determine U_s.

We associate the set $\{-,+\}^n$ with the strict total order $<$ defined as $(s_1, ..., s_n) < (s'_1, ..., s'_n)$ if and only if there exists $i \in \{1, ..., n\}$ such that $s_i = -$, $s'_i = +$ and $s_j = s'_j\ \forall j > i$.

A. Encoding

Let $\{P_s\}_{s\in\{-,+\}^n}$ be a set of 2^n independent copies of the channel P. P_s should not be confused with P^s: P_s is a copy of the channel P and P^s is a polarized channel obtained from P as before.

Define U_{s_1,s_2} for $s_1 \in \{-,+\}^l, s_2 \in \{-,+\}^{n-l}$, $0 \leq l \leq n$, inductively as:

- $U_{\emptyset,s} = U_s$ if $l = 0$, $s \in \{-,+\}^n$.
- $U_{(s_1;-),s_2} = U_{s_1,(s_2;+)} * U_{s_1,(s_2;-)}$ if $l > 0$, $s_1 \in \{-,+\}^{l-1}$, $s_2 \in \{-,+\}^{n-l}$.
- $U_{(s_1;+),s_2} = U_{s_1,(s_2;+)}$ if $l > 0$, $s_1 \in \{-,+\}^{l-1}$, $s_2 \in \{-,+\}^{n-l}$.

We send $U_{s,\emptyset}$ through the channel P_s for all $s \in \{-,+\}^n$. Let Y_s be the output of the channel P_s, and let $Y = \{Y_s\}_{s\in\{-,+\}^n}$. We can prove by induction on l that the channel $U_{s_1,s_2} \to \big(\{Y_s\}_{s \text{ has } s_1 \text{ as a prefix}}, \{U_{s'}\}_{s'<s_2}\big)$ is equivalent to the channel P^{s_2}. In particular, the channel $U_s \to \big(Y, \{U_{s'}\}_{s'<s}\big)$ is equivalent to the channel P^s.

B. Decoding

If $s \notin E_n$ then the receiver knows U_s, there is nothing to decode. Suppose that $s \in E_n$, if we know $\{U_{s'}\}_{s'<s}$ then we can estimate $\text{Proj}_{\mathcal{H}_s}(U_s)$ from $\big(Y, \{U_{s'}\}_{s'<s}\big)$ by the maximum likelihood decoder of $P^s[\mathcal{H}_s]$. After that, we estimate $U_s = f_s(\text{Proj}_{\mathcal{H}_s}(U_s))$. This motivates the following successive cancellation decoder:

- $\hat{U}_s = U_s$ if $s \notin E_n$.
- $\hat{U}_s = \mathcal{D}_s(Y, \{\hat{U}_{s'}\}_{s'<s})$ if $s \in E_n$.

Where $\mathcal{D}_s(Y, \{U_{s'}\}_{s'<s})$ is the estimate of U_s obtained from $(Y, \{U_{s'}\}_{s'<s})$ by the above procedure.

C. Performance of polar codes

If $s \in E_n$, the probability of error in estimating U_s is the probability of error in estimating $\text{Proj}_{\mathcal{H}_s}(U_s)$ using the maximum likelihood decoder, which is upper bounded by

$$|\mathcal{H}_s|.Z(P^s[\mathcal{H}_s]) < |Q|2^{-2^{\beta' n}}.$$

Note that $\mathcal{D}_s(Y, \{U_{s'}\}_{s'<s}) = U_s \ (\forall s \in E_n) \Leftrightarrow \mathcal{D}_s(Y, \{\hat{U}_{s'}\}_{s'<s}) = U_s \ (\forall s \in E_n)$. Therefore, the probability of error of the above successive cancellation decoder is upper bounded by

$$\sum_{s\in E_n} \text{P}\big(\mathcal{D}_s(Y, \{U_{s'}\}_{s'<s}) \neq U_s\big) < |E_n||Q|2^{-2^{\beta' n}} \leq |Q|2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}}.$$

This upper bound was calculated on average over a random choice of the frozen symbols and of the section mappings. Therefore, there exists at least one choice of the frozen symbols and of the section mappings for which the upper bound of the probability of error still holds.

The last thing to discuss is the rate of polar codes. The rate at which we are communicating is $R = \dfrac{1}{2^n}\sum_{s\in E_n} \log|\mathcal{H}_s|$. On the other hand, we have $\big|I(P^s) - \log|\mathcal{H}_s|\big| < \frac{\epsilon}{2}$ for all $s \in E_n$. And since we have $\sum_{s\in\{-,+\}^n} I(P^s) = 2^n I(P)$, we conclude:

$$\begin{aligned} I(P) &= \frac{1}{2^n}\sum_{s\in\{-,+\}^n} I(P^s) \\ &= \frac{1}{2^n}\sum_{s\in E_n} I(P^s) + \frac{1}{2^n}\sum_{s\in E_n^c} I(P^s) \\ &< \frac{1}{2^n}\sum_{s\in E_n}\Big(\log|\mathcal{H}_s| + \frac{\epsilon}{2}\Big) + \frac{1}{2^n}|E_n^c|\log|Q| \\ &< R + \frac{1}{2^n}|E_n|\frac{\epsilon}{2} + \frac{\epsilon}{2\log|Q|}\log|Q| \\ &\leq R + \frac{\epsilon}{2} + \frac{\epsilon}{2} = R + \epsilon. \end{aligned}$$

To this end we have proven the following theorem which is the main result of this paper:

Theorem 5. *Let $P : Q \longrightarrow \mathcal{Y}$ be a channel where the input alphabet has a quasigroup structure. For every $\epsilon > 0$ and for every $0 < \beta < \frac{1}{2}$, there exists a polar code of length N having a rate $R > I(P) - \epsilon$ and a probability of error $P_e < 2^{-N^\beta}$.*

VI. The case of groups

Lemma 14. *Let $(G, *)$ be a group, and let $\mathcal{H}$ be a stable partition of $(G, /^*)$. There exists a normal subgroup of G such that $\mathcal{H}$ is the quotient group of G by H (also denoted by G/H), and $\text{Proj}_{\mathcal{H}}(x) = x \bmod H$ for all $x \in G$.*

Proof: Let H be the element of $\mathcal{H}$ containing the neutral element e of G. For any $H' \in \mathcal{H}$, we have $H' = H'/^*e \subset H'/^*H$. Now because of the stability of $\mathcal{H}$, we have $|H'/^*H| = |H'|$ and so $H'/^*H = H'$ for all $H' \in \mathcal{H}$. This implies that $\mathcal{H}^{/^*} = \mathcal{H}$. Now for any $H_1 \in \mathcal{H} = \mathcal{H}^{/^*}$ and $H_2 \in \mathcal{H}$, there exists $H_3 \in \mathcal{H}$ such that $H_1 = H_3/^*H_2$, and so $H_1 * H_2 = H_3 \in \mathcal{H}$. Therefore, we also have $\mathcal{H}^* = \mathcal{H}$.

Now for any $H' \in \mathcal{H}$, we have $H' = e*H' \subset H*H' \in \mathcal{H}$, $H' = H'*e \subset H'*H \in \mathcal{H}$, and $|H'| = |H*H'| = |H'*H|$, from which we conclude that $H*H' = H'*H = H'$. This implies that $H*H = H$, and $k*H = H*k$ for any $k \in G$. Therefore, H is a normal subgroup of G, and $\mathcal{H}$ is the quotient subgroup of G by H. ∎

By combining the last lemma with theorem 4, we get:

Theorem 6. *Let $P : G \longrightarrow \mathcal{Y}$ be a channel where the input alphabet G has a group structure. P_n converges almost surely to homomorphism channels. Moreover, the convergence is almost surely fast:*

$$\lim_{n\to\infty}\frac{1}{2^n}\Big|\big\{s\in\{-,+\}^n : \exists H \text{ a normal subgroup of } G, \big|I(P^s) - \log|G/H|\big| < \epsilon, \big|I(P^s[H]) - \log|G/H|\big| < \epsilon, Z(P^s[H]) < 2^{-2^{\beta n}}\big\}\Big| = 1,$$

for any $0 < \epsilon < \log 2$, and any $0 < \beta < \frac{1}{2}$. Where $P[H] : G/H \longrightarrow \mathcal{Y}$ is defined as:

$$P[H](y|a) = \frac{1}{|H|}\sum_{\substack{x\in G \\ x \bmod H = a}} P(y|x).$$

VII. Polar codes for arbitrary multiple access channels

In this section, we construct polar codes for an arbitrary multiple access channel, where there is no constraint on the input alphabet sizes: they can be arbitrary, and possibly different from one user to another.

If we have $|\mathcal{X}_k| = p_1^{r_1}p_2^{r_2}...p_{n_k}^{r_{n_k}}$, where p_1, ..., p_{n_k} are prime numbers, we can assume that $\mathcal{X}_k = \mathbb{F}_{p_1}^{r_1}\mathbb{F}_{p_2}^{r_2}...\mathbb{F}_{p_{n_k}}^{r_{n_k}}$, and so we can replace the k^{th} user by $r_1 + r_2 + ... + r_{n_k}$ virtual users having $\mathbb{F}_{p_1}$, $\mathbb{F}_{p_2}$, ..., or $\mathbb{F}_{p_{n_k}}$ as input alphabet respectively. Therefore, we can assume without loss of generality that $\mathcal{X}_k = \mathbb{F}_{q_k}$ for all k, where q_k is a prime number. Let p_1, p_2, ..., p_l be the distinct primes which appear in q_1, ..., q_m, and for each $1 \leq i \leq l$ let m_i be the number of times p_i appears in q_1, ..., q_m.

We adopt two notations to indicate the users and their inputs:

- The first notation is the usual one: we have an index k taking value in $\{1,...,m\}$, and the input of the k^{th} user is denoted by $X_k \in \mathbb{F}_{q_k}$.
- In the second notation, the m_i users having their inputs in $\mathbb{F}_{p_i}$ will be indexed by $(i,1)$, ... , (i,j) , ... , (i,m_i), where $1 \leq i \leq l$ and $1 \leq j \leq m_i$. The input of the $(i,j)^{th}$ user is denoted by $X_{i,j} \in \mathbb{F}_{p_i}$. The vector $(X_{i,1},...,X_{i,m_i}) \in \mathbb{F}_{p_i}^{m_i}$ is denoted by $\vec{X}_i$.

Definition 19. *Let* $P : \prod_{k=1}^{m} \mathbb{F}_{q_k} \to \mathcal{Y}$ *be a discrete* m*-user MAC. We define the two channels* $P^- : \prod_{k=1}^{m} \mathbb{F}_{q_k} \to \mathcal{Y}^2$ *and* $P^+ : \prod_{k=1}^{m} \mathbb{F}_{q_k} \to \mathcal{Y}^2 \times \prod_{k=1}^{m} \mathbb{F}_{q_k}$ *as:*

$$\begin{aligned}
&P^+(y_1,y_2,u_1^1,...,u_m^1|u_1^2,...,u_m^2)\\
&\quad = \frac{1}{q_1...q_m} P(y_1|u_1^1+u_1^2,...,u_m^1+u_m^2)P(y_2|u_1^2,...,u_m^2),\\
&P^-(y_1,y_2|u_1^1,...,u_m^1)\\
&\quad = \sum_{(u_1^2,...,u_m^2)\,\in\,\prod_{k=1}^m \mathbb{F}_{q_k}} P^+(y_1,y_2,u_1^1,...,u_m^1|u_1^2,...,u_m^2),
\end{aligned}$$

where the addition $u_k^1+u_k^2$ *takes place in* $\mathbb{F}_{q_k}$ *if* $u_k^1, u_k^2 \in \mathbb{F}_{q_k}$.

P^- and P^+ can be constructed from two independent copies of P as follows: The k^{th} user chooses independently and uniformly two symbols U_k^1 and U_k^2 in $\mathbb{F}_{q_k}$, then he calculates $X_k^1 = U_k^1 + U_k^2$ and $X_k^2 = U_k^2$, and he finally sends X_k^1 through the first copy of P and X_k^2 through the second copy of P. P^- and P^+ are the channels $U_1^1...U_m^1 \to Y_1Y_2$ and $U_1^2...U_m^2 \to Y_1Y_2U_1^1...U_m^1$ respectively, where Y_1 and Y_2 are the respective outputs of the first and second copy of P respectively.

Note that the transformation $(U_1^1,...,U_m^1,U_1^2,...,U_m^2) \to (X_1^1,...,X_m^1,X_1^2,...,X_m^2)$ is bijective and therefore it induces uniform and independent distributions for $X_1^1,...,X_m^1,X_1^2,...,X_m^2$ which are the inputs of the P channels.

Definition 20. *Let* $\{B_n\}_{n\geq 1}$ *be i.i.d. uniform random variables on* $\{-,+\}$. *We define the* MAC*-valued process* $\{P_n\}_{n\geq 0}$ *by:*

$$\begin{aligned}
&P_0 := P,\\
&P_n := P_{n-1}^{B_n}\ \forall n \geq 1.
\end{aligned}$$

Proposition 1. *The process* $\{I[S](P_n)\}_{n\geq 0}$ *is a bounded super-martingale for all* $S \subset \{1,...,m\}$. *Moreover, it's a bounded martingale if* $S = \{1,...,m\}$.

Proof:

$$\begin{aligned}
2I[S](P) &= I[S](P) + I[S](P)\\
&= I(X^1(S);Y_1X^1(S^c)) + I(X^2(S);Y_2X^2(S^c))\\
&= I(X^1(S)X^2(S);Y_1Y_2X^1(S^c)X^2(S^c))\\
&= I(U^1(S)U^2(S);Y_1Y_2U^1(S^c)U^2(S^c))\\
&= I(U^1(S);Y_1Y_2U^1(S^c)U^2(S^c))\\
&\quad + I(U^2(S);Y_1Y_2U^1(S^c)U^2(S^c)U^1(S))\\
&\geq I(U^1(S);Y_1Y_2U^1(S^c))\\
&\quad + I(U^2(S);Y_1Y_2U_1^1...U_m^1U^2(S^c))\\
&= I[S](P^-) + I[S](P^+).
\end{aligned}$$

Thus, $E\big(I[S](P_{n+1})\big|P_n\big) = \frac{1}{2}I[S](P_n^-) + \frac{1}{2}I[S](P_n^+) \leq I[S](P_n)$, and $I[S](P_n) \leq \sum_{i\in S} \log q_i$ for all $S \subset \{1,...,m\}$, which proves that $\{I[S](P_n)\}_{n\geq 0}$ is a bounded super-martingale. If $S = \{1,...,m\}$, the inequality becomes equality, and $\{I[S](P_n)\}_{n\geq 0}$ is a bounded martingale. ■

From the bounded super-martingale convergence theorem, we deduce that the sequences $\{I[S](P_n)\}_{n\geq 0}$ converge almost surely for all $S \subset \{1,...,m\}$.

Since $\frac{1}{2}(I[S](P^-) + I[S](P^+)) \leq I[S](P)\ \forall S \subset \{1,...,m\}$, then $\frac{1}{2}\mathcal{J}(P-)+\frac{1}{2}\mathcal{J}(P+) \subset \mathcal{J}(P)$, but this subset relation can be strict if one of the inequalities is strict for a certain $S \subset \{1,...,m\}$. Nevertheless, for $S = \{1,...,m\}$, we have $\frac{1}{2}(I(P^-)+I(P^+)) = I(P)$, so at least one point of the dominant face of $\mathcal{J}(P)$ is present in $\frac{1}{2}\mathcal{J}(P-) + \frac{1}{2}\mathcal{J}(P+)$ since the capacity region is a polymatroid. Therefore, the symmetric sum capacity is preserved, but the dominant face might lose some points.

Definition 21. *In order to simplify our notation, we will introduce the notion of generalized matrices:*

- *A generalized matrix* $A = (A_1,...,A_l) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i\times l_i}$ *is a collection of* l *matrices.* $\mathbb{F}_{p_i}^{m_i\times l_i}$ *denotes the set of* $m_i \times l_i$ *matrices with coefficients in* $\mathbb{F}_{p_i}$.
- *If* $l_i = 0$ *in* $A = (A_1,...,A_l) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i\times l_i}$, *we write* $A_i = \emptyset$. *In case* $A_i = \emptyset$ *for all* i, *we write* $A = \emptyset$.
- *A generalized vector* $\vec{x} = (\vec{x}_1,...,\vec{x}_l) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$ *is a collection of* l *vectors.*
- *Addition of generalized vectors is defined naturally as component-wise addition.*
- *The transposition of a generalized matrix is obtained by transposing each matrix of it:* $A^T = (A_1^T,...,A_l^T)$.
- *A generalized matrix operates on a generalized vector component-wise: if* $A \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i\times l_i}$ *and* $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, *then* $\vec{y} = A^T\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{l_i}$ *is defined by* $\vec{y} = (A_1^T\vec{x}_1,...,A_l^T\vec{x}_l)$. *By convention, we have* $\emptyset^T\vec{x}_i = \vec{0}$.

- *A generalized matrix A is said to be full rank if and only if each matrix component of it is full rank.*
- *The rank of a generalized matrix $A \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times l_i}$ is defined by:* $\text{rank}(A) = \sum_{i=1}^{l} \text{rank}(A_i)$.
- *The logarithmic rank of a generalized matrix is defined by:* $\text{lrank}(A) = \sum_{i=1}^{l} \text{rank}(A_i).\log p_i$.
- *If A is a generalized matrix satisfying $A_i \neq \emptyset$ and $A_j = \emptyset$ for all $j \neq i$, we say that A is an ordinary matrix and we identify A and A_i.*

Definition 22. *Let $P : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \to \mathcal{Y}$ be an m-user MAC, let $A \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times l_i}$ be a full rank generalized matrix. We define the* $\text{rank}(A)$*-user MAC $P[A] : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{l_i} \to \mathcal{Y}$ as follows:*

$$P[A](y|\vec{u}) = \frac{1}{\prod_{i=1}^{l} p_i^{m_i - l_i}} \sum_{\substack{\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \\ A^T \vec{x} = \vec{u}}} P(y|\vec{x}).$$

The main result of this section is that, almost surely, P_n becomes a channel where the output is "almost determined by a generalized matrix", and the convergence is almost surely fast:

Theorem 7. *Let $P : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \to \mathcal{Y}$ be an m-user MAC. Then for every $0 < \epsilon < \log 2$, and for every $0 < \beta < \frac{1}{2}$ we have:*

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists A_s \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}}, \; A_s \text{ is full rank}, |I(P^s) - \text{lrank}(A_s)| < \epsilon, \; |I(P^s[A_s]) - \text{lrank}(A_s)| < \epsilon, Z(P^s[A_s]) < 2^{-2^{\beta n}} \Big\} \Big| = 1.$$

Proof: Since $G := \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$ is an abelian group. We can see P as a channel from the Abelian group G to $\mathcal{Y}$. Note that any subgroup of an Abelian group is normal. Therefore, from *theorem 6* we have:

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists H_s \text{ subgroup of } G, \; \big|I(P^s) - \log|G/H_s|\big| < \epsilon, \big|I(P^s[H_s]) - \log|G/H_s|\big| < \epsilon, \; Z(P^s[H_s]) < 2^{-2^{\beta n}} \Big\} \Big| = 1.$$

Let $s \in \{-,+\}^n$ such that that there exists a subgroup H_s of G satisfying:

- $\big|I(P^s) - \log|G/H_s|\big| < \epsilon$.
- $\big|I(P^s[H_s]) - \log|G/H_s|\big| < \epsilon$.
- $Z(P^s[H_s]) < 2^{-2^{\beta n}}$.

From the properties of abelian groups, there exist l integers: $r_{1,s} \leq m_1$, ..., and $r_{l,s} \leq m_l$ such that G/H_s is isomorphic to $\prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_{i,s}}$ (Note that $r_{i,s}$ can be zero). Therefore, there exists a surjective homomorphism $f_s : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \longrightarrow \prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_{i,s}}$, such that for any $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, $f_s(\vec{x})$ can be determined from $\vec{x} \bmod H_s$ and vice versa.

For all $1 \leq i \leq l$, and all $1 \leq j \leq m_i$, define the vector $\vec{e}^{\,i,j} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$ as having all its components as zeros except the $(i,j)^{th}$ component which is equal to 1, then the order of $\vec{e}^{\,i,j}$ is p_i. Let $\vec{y}^{\,i,j} = (\vec{y}_1^{\,i,j}, \vec{y}_2^{\,i,j}, ..., \vec{y}_l^{\,i,j}) = f_s(\vec{e}^{\,i,j}) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_{i,s}}$, if $\vec{y}^{\,i,j} \neq \vec{0}$ then the order of $\vec{y}^{\,i,j}$ must be equal to p_i. If $\vec{y}_{i'}^{\,i,j} \neq \vec{0}$ for a certain $i' \neq i$, then $p_{i'}$ divides the order of $\vec{y}^{\,i,j}$ which is a contradiction. Therefore, we must have $\vec{y}_{i'}^{\,i,j} = \vec{0}$ for all $i' \neq i$.

Now for any $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, we have $\vec{x} = \sum_{i=1}^{l} \sum_{j=1}^{m_i} x_{i,j} \vec{e}^{\,i,j}$, therefore, $f_s(\vec{x}) = \sum_{i=1}^{l} \sum_{j=1}^{m_i} x_{i,j} \vec{y}^{\,i,j}$. Since $\vec{y}_{i'}^{\,i,j} = 0$ for all $i' \neq i$, then $f_s(\vec{x}) = A_s^T \vec{x}$, where $A_s = (A_{1,s}, ..., A_{l,s}) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}}$ is a generalized matrix whose components are given by $A_{i,s} = [\vec{y}_i^{\,i,1} \; \vec{y}_i^{\,i,2} \; ... \; \vec{y}_i^{\,i,m_i}]^T$. A_s is full rank since f_s is surjective. Moreover, we have:

$$\text{lrank}(A_s) = \sum_{i=1}^{l} r_{i,s}.\log p_i = \log\Big(\prod_{i=1}^{l} p_i^{r_{i,s}}\Big) = \log|G/H_s|.$$

Recall that for any $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, $A_s^T \vec{x} = f_s(\vec{x})$ can be determined from $\vec{x} \bmod H_s$ and vice versa, we conclude that $P^s[H_s]$ is equivalent to $P^s[A_s]$. Therefore:

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists A_s \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}}, \; A_s \textit{ is full rank}, |I(P^s) - \textit{lrank}(A_s)| < \epsilon, \; |I(P^s[A_s]) - \textit{lrank}(A_s)| < \epsilon, Z(P^s[A_s]) < 2^{-2^{\beta n}} \Big\} \Big| = 1.$$

■

A. Polar codes construction for MACs

Choose $0 < \epsilon < \log 2$, $0 < \beta < \beta' < \frac{1}{2}$, and let n be an integer such that

- $\Big(\prod_{i=1}^{l} p_i^{m_i}\Big) 2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}}$.
- $\frac{1}{2^n}|E_n| > 1 - \frac{\epsilon}{2\sum_{i=1}^{l} m_i \log p_i}$.

where

$$E_n = \Bigg\{ s \in \{-,+\}^n : \exists A_s \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}}, A_s \text{ is full rank}, \\ |I(P^s) - \text{lrank}(A_s)| < \frac{\epsilon}{2}, |I(P^s[A_s]) - \text{lrank}(A_s)| < \frac{\epsilon}{2}, \\ Z(P^s[A_s]) < 2^{-2^{\beta' n}} \Bigg\}.$$

Such an integer exists due to *theorem* 7.

For each $s \in \{-,+\}^n$, if $s \notin E_n$ set $F(s,i,j) = 1 \ \forall i \in \{1,...,l\} \ \forall j \in \{1,...,m_i\}$, and if $s \in E_n$ choose a generalized matrix $A_s = (A_{1,s},...,A_{l,s})$ that satisfies the conditions in E_n. For each $1 \leq i \leq l$ choose a set of $r_{i,s}$ indices

$$S_{i,s} = \{j_1, ...j_{r_{i,s}}\} \subset \{1,...,m_i\}$$

such that the corresponding rows of $A_{i,s}$ are linearly independent, then set $F(s,i,j) = 1$ if $j \notin S_{i,s}$, and $F(s,i,j) = 0$ if $j \in S_{i,s}$. $F(s,i,j) = 1$ indicates that the user (i,j) is frozen in the channel P^s, i.e., no useful information is being sent.

A polar code is constructed as follows: The user (i,j) sends a symbol $U_{s,i,j}$ through a channel equivalent to P^s. If $F(s,i,j) = 0$, $U_{s,i,j}$ is an information symbol, and if $F(s,i,j) = 1$, $U_{s,i,j}$ is a certain frozen symbol. Since we are free to choose any value for the frozen symbols, we will analyse the performance of the polar code averaged on all the possible choices of the frozen symbols, so we will consider that $U_{s,i,j}$ are independent random variables, uniformly distributed in $\mathbb{F}_{p_i} \ \forall s \in \{-,+\}^n, \forall i \in \{1,...,l\}, \forall j \in \{1,...,m_i\}$. However, the value of $U_{s,i,j}$ will be revealed to the receiver if $F(s,i,j) = 1$, and if $F(s,i,j) = 0$ the receiver has to estimate $U_{s,i,j}$ from the output of the channel.

We associate the set $\{-,+\}^n$ with the strict total order $<$ defined as $s_1...s_n < s'_1...s'_n$ if and only if there exists $i \in \{1,...,n\}$ such that $s_i = -$, $s'_i = +$ and $s_j = s'_j \ \forall j > i$.

1) Encoding: Let $\{P_s\}_{s \in \{-,+\}^n}$ be a set of 2^n independent copies of the channel P. P_s should not be confused with P^s: P_s is a copy of the channel P and P^s is a polarized channel obtained from P as before.

Define $U_{s_1,s_2,i,j}$ for $s_1 \in \{-,+\}^{l'}, s_2 \in \{-,+\}^{n-l'}$, $0 \leq l' \leq n$ inductively as:

- $U_{\emptyset,s,i,j} = U_{s,i,j}$ if $l' = 0$, $s \in \{-,+\}^n$.
- $U_{(s_1;-),s_2,i,j} = U_{s_1,(s_2;+),i,j} + U_{s_1,(s_2;-),i,j}$ if $l' > 0$, $s_1 \in \{-,+\}^{l'-1}$, $s_2 \in \{-,+\}^{n-l'}$.
- $U_{(s_1;+),s_2,i,j} = U_{s_1,(s_2;+),i,j}$ if $l' > 0$, $s_1 \in \{-,+\}^{l'-1}$, $s_2 \in \{-,+\}^{n-l'}$.

The user (i,j) sends $U_{s,\emptyset,i,j}$ through the channel P_s for all $s \in \{-,+\}^n$. Let Y_s be the output of the channel P_s, and let $Y = \{Y_s\}_{s \in \{-,+\}^n}$. We can prove by induction on l' that the channel $\vec{U}_{s_1,s_2} \to \big(\{Y_s\}_{s \text{ has } s_1 \text{ as a prefix}}, \{\vec{U}_{s'}\}_{s' < s_2}\big)$ is equivalent to P^{s_2}. In particular, the channel $\vec{U}_s \to \big(Y, \{\vec{U}_{s'}\}_{s'<s}\big)$ is equivalent to the channel P^s.

2) Decoding: If $s \notin E_n$ then $F(s,i,j) = 1$ for all (i,j), and the receiver knows all $U_{s,i,j}$, there is nothing to decode. Suppose that $s \in E_n$, if we know $\{\vec{U}_{s'}\}_{s'<s}$ then we can estimate $\vec{U}_s$ as follows:

- If $F(s,i,j) = 1$ then we know $U_{s,i,j}$.
- We have $F(s,i,j) = 0$ for $r_{i,s}$ values of j corresponding to $r_{i,s}$ linearly independent rows of $A_{i,s}$. So if we know $A_{i,s}^T \vec{U}_s$, we can recover $U_{s,i,j}$ for the indices j satisfying $F(s,i,j) = 0$.
- Since $A_s^T \vec{U}_s \longrightarrow \big(Y, \{\vec{U}_{s'}\}_{s'<s}\big)$ is equivalent to $P^s[A_s]$, we can estimate $A_s^T \vec{U}_s$ using the maximum likelihood decoder of the channel $P^s[A_s]$.
- Let $\mathcal{D}_s(Y, \{\vec{U}_{s'}\}_{s'<s})$ be the estimate of $\vec{U}_s$ obtained from $(Y, \{\vec{U}_{s'}\}_{s'<s})$ by the above procedure.

This motivates the following successive cancellation decoder:

- $\hat{\vec{U}}_s = \vec{U}_s$ if $s \notin E_n$.
- $\hat{\vec{U}}_s = \mathcal{D}_s(Y, \{\hat{\vec{U}}_{s'}\}_{s'<s})$ if $s \in E_n$.

3) Performance of polar codes: If $s \in E_n$, the probability of error in estimating $A_s^T \vec{U}_s$ using the maximum likelihood decoder is upper bounded by $\Big(\prod_{i=1}^{l} p_i^{r_{i,s}}\Big) Z(P^s[A_s]) < \Big(\prod_{i=1}^{l} p_i^{m_i}\Big) 2^{-2^{\beta' n}}$. Therefore, the probability of error in estimating $\vec{U}_s$ from $(Y, \{\vec{U}_{s'}\}_{s'<s})$ is upper bounded by $\Big(\prod_{i=1}^{l} p_i^{m_i}\Big) 2^{-2^{\beta' n}}$ when $s \in E_n$.

Note that $\mathcal{D}_s(Y, \{\vec{U}_{s'}\}_{s'<s}) = \vec{U}_s, \ (\forall s \in E_n) \Leftrightarrow \mathcal{D}_s(Y, \{\hat{\vec{U}}_{s'}\}_{s'<s}) = \vec{U}_s \ (\forall s \in E_n)$, so the probability of error of the above successive cancellation decoder is upper bounded by

$$\sum_{s \in E_n} \mathrm{P}\big(\mathcal{D}_s(Y, \{\vec{U}_{s'}\}_{s'<s}) \neq \vec{U}_s\big) \\ < |E_n| \Big(\prod_{i=1}^{l} p_i^{m_i}\Big) 2^{-2^{\beta' n}} \leq \Big(\prod_{i=1}^{l} p_i^{m_i}\Big) 2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}}.$$

The above upper bound was calculated on average over a random choice of the frozen symbols. Therefore, there is at least one choice of the frozen symbols for which the upper bound of the probability of error still holds.

The last thing to discuss is the rate vector of polar codes. The rate at which the user (i,j) is communicating is $R_{i,j} =$

$\frac{1}{2^n}\sum_{s\in E_n}\big(1-F(s,i,j)\big)\log p_i$, the sum rate is:

$$\begin{aligned}R&=\sum_{1\le i\le l}\sum_{1\le j\le m_i}R_{i,j}\\&=\frac{1}{2^n}\sum_{1\le i\le l}\sum_{1\le j\le m_i}\sum_{s\in E_n}\big(1-F(s,i,j)\big)\log p_i\\&=\frac{1}{2^n}\sum_{s\in E_n}\sum_{1\le i\le l}r_{i,s}\log p_i=\frac{1}{2^n}\sum_{s\in E_n}\text{lrank}(A_s).\end{aligned}$$

We have $|I(P^s)-\text{lrank}(A_s)|<\frac{\epsilon}{2}$ and $I(P^s)<\text{lrank}(A_s)+\frac{\epsilon}{2}$ for all $s\in E_n$. And since we have $\sum_{s\in\{-,+\}^n}I(P^s)=2^nI(P)$ we conclude:

$$\begin{aligned}I(P)&=\frac{1}{2^n}\sum_{s\in\{-,+\}^n}I(P^s)\\&=\frac{1}{2^n}\sum_{s\in E_n}I(P^s)+\frac{1}{2^n}\sum_{s\in E_n^c}I(P^s)\\&<\frac{1}{2^n}\sum_{s\in E_n}\Big(\text{lrank}(A_s)+\frac{\epsilon}{2}\Big)+\frac{1}{2^n}|E_n^c|\sum_{i=1}^{l}m_i\log p_i\\&<R+\frac{1}{2^n}|E_n|\frac{\epsilon}{2}+\frac{\epsilon}{2\sum_{i=1}^{l}m_i\log p_i}\sum_{i=1}^{l}m_i\log p_i\\&\le R+\frac{\epsilon}{2}+\frac{\epsilon}{2}=R+\epsilon.\end{aligned}$$

To this end we have proven the following theorem which is the main result of this subsection:

Theorem 8. *Let $P:\prod_{i=1}^{l}\mathbb{F}_{p_i}^{m_i}\to\mathcal{Y}$ be an m-user MAC. For every $\epsilon>0$ and for every $0<\beta<\frac{1}{2}$, there exists a polar code of length N having a sum rate $R>I(P)-\epsilon$ and a probability of error $P_e<2^{-N^\beta}$.*

A final note to report is that by changing our choice of the indices in $S_{i,s}$, the rate vector of the polar code moves at a distance of at most ϵ along the dominant face of the capacity region achievable by polar codes. However, the dominant face of the initial capacity region can be strictly bigger than the dominant face achievable by polar codes, in such case we say that we have a loss in the dominant face.

VIII. Case study

In this section, we are interested in studying the problem of loss in the capacity region by polarization in the special case of channels which are combination of deterministic linear channels. For simplicity, we will consider MAC channels where the input alphabet size is a prime number q and which is the same for all the users. Moreover we will use the base-q logarithm in the expression of the mutual informations and entropies.

Definition 23. *An m-user MAC P is said to be a combination of l linear channels, if there are l matrices $A_1,...,A_l$, ($A_k\in\mathbb{F}_q^{m\times m_k}$) such that P is equivalent to the channel $P_{lin}:\mathbb{F}_q^m\to\bigcup_{k=1}^{l}\{k\}\times\mathbb{F}_q^{m_k}$ defined by:*

$$P_{lin}(k,\vec{y}|\vec{x})=\begin{cases}p_k & \text{if } A_k^T\vec{x}=\vec{y}\\0 & \text{otherwise}\end{cases}$$

where $\sum_{k=1}^{l}p_k=1$ and $p_k\neq0$ $\forall k$. The channel P_{lin} is denoted by $P_{lin}=\sum_{k=1}^{l}p_k\mathcal{C}_{A_k}$.

The channel P_{lin} can be seen as a box where we have a collection of matrices. At each channel use, a matrix A_k from the box is chosen randomly according to the probabilities p_k, and the output of the channel is $A_k^T\vec{x}$, together with the index k (so the receiver knows which matrix has been used).

A. Characterizing non-losing channels

In the case of channels that are combination of linear channels, we are interested in finding the channels whose capacity region is preserved upon the polarization process.

Proposition 2. *If $\{A_k,A'_k:1\le k\le l\}$ is a set of matrices such that $\text{span}(A_k)=\text{span}(A'_k)$ $\forall k$, then the two channels $P=\sum_{k=1}^{l}p_k\mathcal{C}_{A_k}$ and $P'=\sum_{k=1}^{l}p_k\mathcal{C}_{A'_k}$ are equivalent.*

Proof: If $\text{span}(A_k)=\text{span}(A'_k)$, we can determine $A_k^T\vec{x}$ from ${A'_k}^T\vec{x}$ and vice versa. Therefore, from the output of P, we can deterministically obtain the output of P' and vice versa. In this sense, P and P' are equivalent, and have the same capacity region. ∎

Notation 4. *Motivated by the above proposition, we will write $P\equiv\sum_{k=1}^{l}p_k\mathcal{C}_{V_k}$ (where $\{V_k\}_{1\le k\le l}$ is a set of l subspaces of $\mathbb{F}_q^m$), whenever P is equivalent to $\sum_{k=1}^{l}p_k\mathcal{C}_{A_k}$ and $\text{span}(A_k)=V_k$.*

Proposition 3. *If $P\equiv\sum_{k=1}^{l}p_k\mathcal{C}_{V_k}$, then $I[S](P)=\sum_{k=1}^{l}p_k\dim\big(\text{proj}_S(V_k)\big)$ for all $S\subset\{1,...,m\}$. Where proj_S denotes the canonical projection on $\mathbb{F}_q^S$ defined by $\text{proj}_S(\vec{x})=\text{proj}_S(x_1,...,x_m)=(x_{i_1},...,x_{i_{|S|}})$ for $\vec{x}=(x_1,...,x_m)\in\mathbb{F}_q^m$ and $S=\{i_1,...,i_{|S|}\}$.*

Proof: Let $X_1,...,X_m$ be the input to the channel $\sum_{k=1}^{l}p_k\mathcal{C}_{A_k}$ (where A_k spans V_k), and let $K,\vec{Y}$ be the output

of it. We have:

$$\begin{aligned}
&H\big(X(S)|K,\vec{Y},X(S^c)\big)\\
&\quad=\sum_{k,\vec{y}}\mathrm{P}_{K,\vec{Y}}(k,\vec{y})H\big(X(S)|k,\vec{y},X(S^c)\big)\\
&\quad=\sum_{k,\vec{y}}\sum_{\vec{x}}\mathrm{P}_{K,\vec{Y}|\vec{X}}(k,\vec{y}|\vec{x})\mathrm{P}_{\vec{X}}(\vec{x})H\big(X(S)|k,\vec{y},X(S^c)\big)\\
&\quad=\sum_{k,\vec{y}}\sum_{\substack{\vec{x},\\ A_k^T\vec{x}=\vec{y}}}p_k\mathrm{P}_{\vec{X}}(\vec{x})H\big(X(S)|k,\vec{y},X(S^c)\big)\\
&\quad=\sum_{k}p_kH\big(X(S)|A_k^T\vec{X},X(S^c)\big)\\
&\quad=\sum_{k}p_kH\big(X(S)|A_k(S)^T\vec{X}(S),X(S^c)\big)\\
&\quad=\sum_{k}p_kH\big(X(S)|A_k(S)^T\vec{X}(S)\big).
\end{aligned}$$

The last equality follows from the fact that $X(S)$ and $X(S^c)$ are independent. $A_k(S)$ is obtained from A_k by taking the rows corresponding to S. For a given value of $A_k(S)^T\vec{X}(S)$, we have q^{d_k} possible values of $\vec{X}(S)$ with equal probabilities, where d_k is the dimension of the null space of the mapping $\vec{X}(S)\to A_k(S)^T\vec{X}(S)$, so we have $H\big(X(S)|A_k(S)^T\vec{X}(S)\big)=d_k$.

On the other hand, $|S|-H\big(X(S)|A_k(S)^T\vec{X}(S)\big)=|S|-d_k$ is the dimension of the range space of the the mapping $\vec{X}(S)\to A_k(S)^T\vec{X}(S)$, which is also equal to the rank of $A_k(S)^T$. Therefore, we have:

$$\begin{aligned}
&|S|-H\big(X(S)|A_k(S)^T\vec{X}(S)\big)\\
&\quad=\mathrm{rank}(A_k(S)^T)=\mathrm{rank}\big(A_k(S)\big)=\dim\Big(\mathrm{span}\big(A_k(S)\big)\Big)\\
&\quad=\dim\Big(\mathrm{proj}_S\big(\mathrm{span}(A_k)\big)\Big)=\dim\big(\mathrm{proj}_S(V_k)\big).
\end{aligned}$$

We conclude:

$$\begin{aligned}
&I(X(S);K,Y,X(S^c))\\
&\qquad=H(X(S))-H(X(S)|K,Y,X(S^c))\\
&\qquad=|S|-\sum_k p_kH(X(S)|A_k(S)^T\vec{X}(S))\\
&\qquad=\sum_k p_k\big(|S|-H(X(S)|A_k(S)^T\vec{X}(S))\big)\\
&\qquad=\sum_k p_k(|S|-d_k)=\sum_k p_k\dim(\mathrm{proj}_S(V_k)).
\end{aligned}$$

■

Proposition 4. *If* $P\equiv\sum_{k=1}^{l}p_k\mathcal{C}_{V_k}$ *then:*

- $P^-\equiv\sum_{k_1=1}^{l}\sum_{k_2=1}^{l}p_{k_1}p_{k_2}\mathcal{C}_{V_{k_1}\cap V_{k_2}}$.
- $P^+\equiv\sum_{k_1=1}^{l}\sum_{k_2=1}^{l}p_{k_1}p_{k_2}\mathcal{C}_{V_{k_1}+V_{k_2}}$.

Proof: Suppose without lost of generality that $P=\sum_{k=1}^{l}p_k\mathcal{C}_{A_k}$ where A_k spans V_k. Let $\vec{U}_1$ be an arbitrarily distributed random vector in $\mathbb{F}_q^m$ (not necessarily uniform), let $\vec{U}_2$ be a uniformly distributed random vector in $\mathbb{F}_q^m$ and independent from $\vec{U}_1$. Let $\vec{X}_1=\vec{U}_1+\vec{U}_2$ and $\vec{X}_2=\vec{U}_2$. Let $(K_1,A_{K_1}^T\vec{X}_1)$ and $(K_2,A_{K_2}^T\vec{X}_2)$ be the output of P when the input is X_1 and X_2 respectively. Then the channel $\vec{U}_1\to(K_1,A_{K_1}^T\vec{X}_1,K_2,A_{K_2}^T\vec{X}_2)$ is equivalent to P^- with $\vec{U}_1$ as input. We did not put any constraint on the distribution of $\vec{U}_1$ (such as saying that $\vec{U}_1$ is uniform) because in general, the model of a channel is characterized by its conditional probabilities and no assumption is made on the input probabilities.

Fix $K_1=k_1$ and $K_2=k_2$, let $A_{k_1\wedge k_2},B_{k_1}$ and B_{k_2} be three matrices chosen such that $A_{k_1\wedge k_2}$ spans $V_{k_1}\cap V_{k_2}$, $A_{k_1}:=[A_{k_1\wedge k_2}\ B_{k_1}]$ spans V_{k_1}, $A_{k_2}:=[A_{k_1\wedge k_2}\ B_{k_2}]$ spans V_{k_2}, and the columns of $[A_{k_1\wedge k_2}\ B_{k_1}\ B_{k_2}]$ are linearly independent. Then knowing $A_{k_1}^T\vec{X}_1$ and $A_{k_2}^T\vec{X}_2$ is equivalent to knowing $A_{k_1\wedge k_2}^T(\vec{U}_1+\vec{U}_2)$, $B_{k_1}^T(\vec{U}_1+\vec{U}_2)$, $A_{k_1\wedge k_2}^T\vec{U}_2$ and $B_{k_2}^T\vec{U}_2$, which is equivalent to knowing $\vec{T}^1_{k_1,k_2}=A_{k_1\wedge k_2}^T\vec{U}_1$, $\vec{T}^2_{k_1,k_2}=B_{k_1}^T(\vec{U}_1+\vec{U}_2)$ and $\vec{T}^3_{k_1,k_2}=[A_{k_1\wedge k_2}\ B_{k_2}]^T\vec{U}_2$. We conclude that P^- is equivalent to the channel:

$$\vec{U}_1\to\big(K_1,K_2,\vec{T}^1_{K_1,K_2},\vec{T}^2_{K_1,K_2},\vec{T}^3_{K_1,K_2}\big).$$

Conditioned on $(K_1,K_2,\vec{T}^1_{K_1,K_2})$ we have $[B_{K_1}\ A_{K_1\wedge K_2}\ B_{K_2}]^T\vec{U}_2$ is uniform (since the matrix $[B_{K_1}\ A_{K_1\wedge K_2}\ B_{K_2}]$ is full rank) and independent from $\vec{U}_1$, so $[A_{K_1\wedge K_2}\ B_{K_2}]^T\vec{U}_2$ is independent from $(B_{K_1}^T\vec{U}_2,\vec{U}_1)$, which implies that $[A_{K_1\wedge K_2}\ B_{K_2}]^T\vec{U}_2$ is independent from $\big(B_{K_1}^T(\vec{U}_1+\vec{U}_2),\vec{U}_1\big)$. Also conditioned on $(K_1,K_2,\vec{T}^1_{K_1,K_2})$, $B_{K_1}^T\vec{U}_2$ is uniform and independent from $\vec{U}_1$, which implies that $\vec{U}_1$ is independent from $B_{K_1}^T(\vec{U}_1+\vec{U}_2)$, and this is because the columns of B_{K_1} and $A_{K_1\wedge K_2}$ are linearly independent. We conclude that conditioned on $(K_1,K_2,\vec{T}^1_{K_1,K_2})$, $\vec{U}_1$ is independent from $\big(\vec{T}^2_{K_1,K_2},\vec{T}^3_{K_1,K_2}\big)$. Therefore, $\big(K_1,K_2,\vec{T}^1_{K_1,K_2}\big)=\big(K_1,K_2,A_{k_1\wedge k_2}^T\vec{U}_1\big)$ form sufficient statistics. We conclude that P^- is equivalent to the channel:

$$\vec{U}_1\to=\big(K_1,K_2,A_{k_1\wedge k_2}^T\vec{U}_1\big).$$

And since $\mathrm{P}(K_1=k_1,K_2=k_2)=p_{k_1}p_{k_2}$, and $A_{k_1\wedge k_2}$ spans $V_{k_1}\cap V_{k_2}$ we conclude that $P^-\equiv\sum_{k_1=1}^{l}\sum_{k_2=1}^{l}p_{k_1}p_{k_2}\mathcal{C}_{V_{k_1}\cap V_{k_2}}$.

Now let $\vec{U}_2$ be arbitrarily distributed in $\mathbb{F}_q^m$ (not necessarily uniform) and $\vec{U}_1$ be a uniformly distributed random vector in $\mathbb{F}_q^m$ independent from $\vec{U}_2$. Let $\vec{X}_1=\vec{U}_1+\vec{U}_2$ and $\vec{X}_2=\vec{U}_2$. Let $(K_1,A_{K_1}^T\vec{X}_1)$ and $(K_2,A_{K_2}^T\vec{X}_2)$ be the output of P when the input is X_1 and X_2 respectively. Then the channel $\vec{U}_2\to(K_1,A_{K_1}^T\vec{X}_1,K_2,A_{K_2}^T\vec{X}_2,\vec{U}_1)$ is equivalent to P^+ with $\vec{U}_2$ as input. Note that the uniform distribution constraint is now on

$\vec{U}_1$ and no constraint is put on the distribution of $\vec{U}_2$, since now $\vec{U}_2$ is the input to the channel P^+.

Knowing $A_{K_1}^T\vec{X}_1$, $A_{K_2}^T\vec{X}_2$ and $\vec{U}_1$ is equivalent to knowing $A_{K_1}^T(\vec{U}_1+\vec{U}_2)$, $A_{K_2}^T\vec{U}_2$ and $\vec{U}_1$, which is equivalent to knowing $A_{K_1}^T\vec{U}_2$, $A_{K_2}^T\vec{U}_2$ and $\vec{U}_1$. So P^+ is equivalent to the channel:

$$\vec{U_2} \rightarrow \big(K_1, K_2, [A_{k_1}\ A_{k_2}]^T\vec{U}_2, \vec{U}_1\big).$$

And since $\vec{U}_1$ is independent from $\vec{U}_2$, the above channel (and hence $P+$) is equivalent to the channel:

$$\vec{U_2} \rightarrow \big(K_1, K_2, [A_{k_1}\ A_{k_2}]^T\vec{U}_2\big).$$

We also have $\mathrm{P}(K_1 = k_1, K_2 = k_2) = p_{k_1}p_{k_2}$, and $[\underset{l}{A_{k_1}}\ \underset{l}{A_{k_2}}]$ spans $V_{k_1} + V_{k_2}$. We conclude that $P^+ \equiv \sum_{k_1=1}^{l}\sum_{k_2=1}^{l} p_{k_1}p_{k_2}\mathcal{C}_{V_{k_1}+V_{k_2}}$. ∎

Lemma 15. *Let $P \equiv \sum_{k=1}^{l} p_k\mathcal{C}_{V_k}$ and $S \subset \{1, ..., m\}$, then*

$$\frac{1}{2}\big(I[S](P^-) + I[S](P^+)\big) = I[S](P) \Leftrightarrow \Big(\forall(k_1, k_2);\ \mathrm{proj}_S(V_{k_1} \cap V_{k_2}) = \mathrm{proj}_S(V_{k_1}) \cap \mathrm{proj}_S(V_{k_2})\Big).$$

Proof: We know that if V and V' are two subspaces of $\mathbb{F}_q^m$, then $\mathrm{proj}_S(V \cap V') \subset \mathrm{proj}_S(V) \cap \mathrm{proj}_S(V')$ and $\mathrm{proj}_S(V + V') = \mathrm{proj}_S(V) + \mathrm{proj}_S(V')$, which implies that:

- $\dim\big(\mathrm{proj}_S(V \cap V')\big) \leq \dim\big(\mathrm{proj}_S(V) \cap \mathrm{proj}_S(V')\big)$.
- $\dim\big(\mathrm{proj}_S(V + V')\big) = \dim\big(\mathrm{proj}_S(V) + \mathrm{proj}_S(V')\big)$.

We conclude:

$$\begin{aligned}\dim\big(\mathrm{proj}_S(V \cap V')\big) &+ \dim\big(\mathrm{proj}_S(V + V')\big)\\ &\leq \dim\big(\mathrm{proj}_S(V) \cap \mathrm{proj}_S(V')\big)\\ &\qquad + \dim\big(\mathrm{proj}_S(V) + \mathrm{proj}_S(V')\big)\\ &= \dim\big(\mathrm{proj}_S(V)\big) + \dim\big(\mathrm{proj}_S(V')\big).\end{aligned}$$

Therefore:

$$\begin{aligned}&\frac{1}{2}\big(I[S](P^-) + I[S](P^+)\big)\\ &= \frac{1}{2}\sum_{k_1=1}^{l}\sum_{k_2=1}^{l} p_{k_1}p_{k_2}\dim\big(\mathrm{proj}_S(V_{k_1} \cap V_{k_2})\big)\\ &\quad + \frac{1}{2}\sum_{k_1=1}^{l}\sum_{k_2=1}^{l} p_{k_1}p_{k_2}\dim\big(\mathrm{proj}_S(V_{k_1} + V_{k_2})\big)\Big)\\ &= \frac{1}{2}\Big(\sum_{k_1=1}^{l}\sum_{k_2=1}^{l} p_{k_1}p_{k_2}\Big(\dim\big(\mathrm{proj}_S(V_{k_1} \cap V_{k_2})\big)\\ &\qquad + \dim\big(\mathrm{proj}_S(V_{k_1} + V_{k_2})\big)\Big)\Big)\\ &\leq \frac{1}{2}\Big(\sum_{k_1=1}^{l}\sum_{k_2=1}^{l} p_{k_1}p_{k_2}\Big(\dim\big(\mathrm{proj}_S(V_{k_1})\big)\\ &\qquad + \dim\big(\mathrm{proj}_S(V_{k_2})\big)\Big)\Big)\\ &= \frac{1}{2}\Big(\sum_{k_1=1}^{l} p_{k_1}\dim\big(\mathrm{proj}_S(V_{k_1})\big) + \sum_{k_2=1}^{l} p_{k_2}\dim\big(\mathrm{proj}_S(V_{k_2})\big)\Big)\\ &= \frac{1}{2}(I[S](P) + I[S](P)) = I[S](P).\end{aligned}$$

So if we have $\mathrm{proj}_S(V_{k_1} \cap V_{k_2}) \subsetneq \mathrm{proj}_S(V_{k_1}) \cap \mathrm{proj}_S(V_{k_2})$ for some k_1, k_2, then we have $\dim\big(\mathrm{proj}_S(V_{k_1} \cap V_{k_2})\big) < \dim\big(\mathrm{proj}_S(V_{k_1}) \cap \mathrm{proj}_S(V_{k_2})\big)$, and the above inequality of mutual information will be strict. We conclude that:

$$\frac{1}{2}\big(I[S](P^-) + I[S](P^+)\big) = I[S](P) \Leftrightarrow \Big(\forall(k_1, k_2);\ \mathrm{proj}_S(V_{k_1} \cap V_{k_2}) = \mathrm{proj}_S(V_{k_1})\mathrm{proj}_S(V_{k_2})\Big).$$

∎

Definition 24. *Let $\mathcal{V}$ be a set of subspaces of $\mathbb{F}_q^m$, we define the closure of $\mathcal{V}$, $cl(\mathcal{V})$, as being the minimal set of subspaces of $\mathbb{F}_q^m$ closed under the two operations $\cap$ and $+$, and including $\mathcal{V}$. We say that the set $\mathcal{V}$ is consistent with respect to $S \subset \{1, ..., m\}$ if and only if it satisfies the following property:*

$$\Big(\forall(V_1, V_2) \in cl(\mathcal{V});\ \mathrm{proj}_S(V_{k_1} \cap V_{k_2}) = \mathrm{proj}_S(V_{k_1}) \cap \mathrm{proj}_S(V_{k_2})\Big).$$

Corollary 1. *If $\mathcal{V} = \{V_k : 1 \leq k \leq l\}$. $I[S](P)$ is preserved upon the polarization process if and only if $\mathcal{V}$ is consistent with respect to S.*

Proof: Upon the polarization process, we are performing successively the $\cap$ and $+$ operators, which means that we'll reach the closure of $\mathcal{V}$ after a finite number of steps. So $I[S](P)$ is preserved if and only if the above lemma applies to $cl(\mathcal{V})$. ∎

The above corollary gives a characterization for a combination of linear channels to preserve $I[S](P)$. However, this characterization involves using the closure operator. The next proposition gives a sufficient condition that uses only the initial

configuration of subspaces $\mathcal{V}$. This proposition gives a certain "geometric" view of what the subspaces should look like if we don't want to lose.

Proposition 5. *If there exists a subspace V_S of dimension $|S|$ whose projection on S is $\mathbb{F}_q^S$ (i.e. $\text{proj}_S(V_S) = \mathbb{F}_q^S$), such that for every $V \in \mathcal{V}$ we have $\text{proj}_S(V_S \cap V) = \text{proj}_S(V)$, then $I[S](P)$ is preserved upon the polarization process. In other words, if every subspace in $\mathcal{V}$ passes through V_S "orthogonally" to S, then $I[S](P)$ is preserved upon the polarization process.*

Proof: Let V_S be a subspace satisfying the hypothesis, then it satisfies also the hypothesis if we replace $\mathcal{V}$ by it's closure: If V_1 and V_2 are two arbitrary subspaces satisfying

$$\text{proj}_S(V_S \cap V_1) = \text{proj}_S(V_1) \text{ and } \text{proj}_S(V_S \cap V_2) = \text{proj}_S(V_2)$$

then $\text{proj}_S(V_1) \subset \text{proj}_S\big(V_S \cap (V_1 + V_2)\big)$ and $\text{proj}_S(V_2) \subset \text{proj}_S\big(V_S \cap (V_1 + V_2)\big)$, which implies $\text{proj}_S(V_1 + V_2) = \text{proj}_S(V_1) + \text{proj}_S(V_2) \subset \text{proj}_S\big(V_S \cap (V_1 + V_2)\big)$. Therefore, $\text{proj}_S\big(V_S \cap (V_1 + V_2)\big) = \text{proj}_S(V_1 + V_2)$ since the inverse inclusion is trivial.

Now let $\vec{x} \in \text{proj}_S(V_1) \cap \text{proj}_S(V_2)$, then $\vec{x} \in \text{proj}_S(V_1) = \text{proj}_S(V_1 \cap V_S)$ and similarly $\vec{x} \in \text{proj}_S(V_2 \cap V_S)$ which implies that there are two vectors $\vec{x_1} \in V_1 \cap V_S$ and $\vec{x_2} \in V_2 \cap V_S$ such that $\vec{x} = \text{proj}_S(\vec{x_1}) = \text{proj}_S(\vec{x_2})$. And since $\text{proj}_S(V_S) = \mathbb{F}_q^S$ and $\dim(V_S) = |S|$, then the mapping $\text{proj}_S : V_S \to \mathbb{F}_q^S$ is invertible and so $\vec{x_1} = \vec{x_2}$ which implies that $\vec{x} \in \text{proj}_S(V_1 \cap V_2 \cap V_S)$. Thus $\text{proj}_S(V_1) \cap \text{proj}_S(V_2) \subset \text{proj}_S(V_1 \cap V_2) \subset \text{proj}_S(V_1 \cap V_2 \cap V_S)$. We conclude that $\text{proj}_S(V_1) \cap \text{proj}_S(V_2) = \text{proj}_S(V_1 \cap V_2) = \text{proj}_S(V_1 \cap V_2 \cap V_S)$ since the inverse inclusions are trivial.

We conclude that the set of subspaces V satisfying $\text{proj}_S(V \cap V_S) = \text{proj}_S(V)$ is closed under the two operators $\cap$ and $+$. And since $\mathcal{V}$ is a subset of this set, $cl(\mathcal{V})$ is a subset as well. Now let $V_1, V_2 \in cl(\mathcal{V})$, then $\text{proj}_S(V_S \cap V_1) = \text{proj}_S(V_1)$ and $\text{proj}_S(V_S \cap V_2) = \text{proj}_S(V_2)$. Then $\text{proj}_S(V_1) \cap \text{proj}_S(V_2) = \text{proj}_S(V_1 \cap V_2)$ as we have seen in the previous paragraph. We conclude that $\mathcal{V}$ is consistent with respect to S and so $I[S](P)$ is preserved. ∎

Conjecture 1. *The condition in proposition 7 is necessary.*

B. Maximal loss in the dominant face

After characterizing the non-losing channels, we are now interested in studying the amount of loss in the capacity region. In order to simplify the problem, we only study it in the case of binary input 2-user MAC since we can easily generalize for the general case.

Since we only have 5 subspaces of $\mathbb{F}_2^2$, we write $P \equiv \sum_{k=0}^{4} p_k \mathcal{C}_{V_k}$ (here p_k is allowed to be zero), where V_0, ..., V_4 are the 5 possible subspaces of $\mathbb{F}_2^2$:

- $V_0 = \{(0,0)\}$.
- $V_1 = \{(0,0),(1,0)\}$.
- $V_2 = \{(0,0),(0,1)\}$.
- $V_3 = \{(0,0),(1,1)\}$.
- $V_4 = \{(0,0),(1,0),(0,1),(1,1)\}$.

We have $I[\{1\}](P) = p_1 + p_3 + p_4$, $I[\{2\}](P) = p_2 + p_3 + p_4$ and $I(P) = I[\{1,2\}](P) = p_1 + p_2 + p_3 + 2p_4$.

Definition 25. *Let $P \equiv \sum_{k=0}^{4} p_k \mathcal{C}_{V_k}$ and $s \in \{-,+\}^n$, we write p_k^s to denote the component of V_k in P^s, i.e. we have $P^s \equiv \sum_{k=0}^{4} p_k^s \mathcal{C}_{V_k}$.*

We denote the average of p_k^s on all possible $s \in \{-,+\}^n$ by $p_k^{(n)}$. i.e. $p_k^{(n)} = \frac{1}{2^n} \sum_{s \in \{-,+\}^n} p_k^s$. $p_k^{(\infty)}$ is the limit of $p_k^{(n)}$ as n tends to infinity.

We denote the average of $I[\{1\}](P^s)$ (resp. $I[\{2\}](P^s)$ and $I(P^s)$) on all possible $s \in \{-,+\}^n$ by $I_1^{(n)}$ (resp. $I_2^{(n)}$ and $I^{(n)}$). We have $I_1^{(n)} = p_1^{(n)} + p_3^{(n)} + p_4^{(n)}$, $I_2^{(n)} = p_2^{(n)} + p_3^{(n)} + p_4^{(n)}$ and $I^{(n)} = p_1^{(n)} + p_2^{(n)} + p_3^{(n)} + 2p_4^{(n)}$. If n tends to infinity we get $I_1^{(\infty)} = p_1^{(\infty)} + p_3^{(\infty)} + p_4^{(\infty)}$, $I_2^{(\infty)} = p_2^{(\infty)} + p_3^{(\infty)} + p_4^{(\infty)}$ and $I^{(\infty)} = p_1^{(\infty)} + p_2^{(\infty)} + p_3^{(\infty)} + 2p_4^{(\infty)}$.

Definition 26. *We say that we have maximal loss in the dominant face in the polarization process, if the dominant face of the capacity region converges to a single point.*

Remark 5. *The symmetric capacity region after n polarization steps is the average of the symmetric capacity regions of all the channels P^s obtained after n polarization steps ($s \in \{-,+\}^n$). Therefore, this capacity region is given by:*

$$\mathcal{J}(P^{(n)}) := \Big\{(R_1,R_2) :\ 0 \le R_1 \le I_1^{(n)},\ 0 \le R_2 \le I_2^{(n)},\ 0 \le R_1 + R_2 \le I^{(n)}\Big\}.$$

The above capacity region converges to the "final capacity region":

$$\mathcal{J}(P^{(\infty)}) := \Big\{(R_1,R_2) :\ 0 \le R_1 \le I_1^{(\infty)},\ 0 \le R_2 \le I_2^{(\infty)},\ 0 \le R_1 + R_2 \le I^{(\infty)}\Big\}.$$

The dominant face converges to a single point if and only if $I^{(\infty)} = I_1^{(\infty)} + I_2^{(\infty)}$, which is equivalent to $p_1^{(\infty)} + p_2^{(\infty)} + p_3^{(\infty)} + 2p_4^{(\infty)} = p_1^{(\infty)} + p_2^{(\infty)} + 2p_3^{(\infty)} + 2p_4^{(\infty)}$. We conclude that we have maximal loss in the dominant face if and only if $p_3^{(\infty)} = 0$.

Lemma 16. *The order of p_1, p_2 and p_3 remains the same upon the polarization process. e.g. if $p_1 < p_3 < p_2$ then $p_1^s < p_3^s < p_2^s$, and if $p_2 = p_3 < p_1$ then $p_2^s = p_3^s < p_1^s$ for all $s \in \{-,+\}^n$.*

Proof: We have $P^- = \sum_{k=0}^{4} \sum_{k'=0}^{4} p_k p_{k'} \mathcal{C}_{V_k \cap V_{k'}}$ and $P^+ =$

$\sum_{k=0}^{4}\sum_{k'=0}^{4} p_k p_{k'} \mathcal{C}_{V_k+V_{k'}}$. Therefore, we have:

$$
\begin{aligned}
p_0^- &= p_0^2 + 2p_0(p_1+p_2+p_3+p_4) + 2(p_1p_2+p_2p_3+p_1p_3),\\
p_1^- &= p_1^2 + 2p_1p_4,\\
p_2^- &= p_2^2 + 2p_2p_4,\\
p_3^- &= p_3^2 + 2p_3p_4,\\
p_4^- &= p_4^2,
\end{aligned}
$$

$$
\begin{aligned}
p_0^+ &= p_0^2,\\
p_1^+ &= p_1^2 + 2p_1p_0,\\
p_2^+ &= p_2^2 + 2p_2p_0,\\
p_3^+ &= p_3^2 + 2p_3p_0,\\
p_4^+ &= p_4^2 + 2p_4(p_1+p_2+p_3+p_4) + 2(p_1p_2+p_2p_3+p_1p_3).
\end{aligned}
$$

We can easily see that the order of p_1^-,p_2^- and p_3^- is the same as that of p_1, p_2 and p_3. This is also true for p_1^+,p_2^+ and p_3^+. By using a simple induction on s, we conclude that the order of p_1^s, p_2^s and p_3^s is the same as that of p_1, p_2 and p_3 for all $s \in \{-,+\}^n$. ■

Lemma 17. *For $k \in \{1,2,3\}$, if $\exists k' \in \{1,2,3\} \setminus \{k\}$ such that $p_k \leq p_{k'}$ then*

$$p_k^{(\infty)} = \lim_{l\to\infty} \frac{1}{2^n} \sum_{s\in\{-,+\}^n} p_k^s = 0.$$

In other words, the component of V_k is killed by that of $V_{k'}$.

Proof: We know from *theorem 7* that the channel P^s converges almost surely to a deterministic linear channel as n tends to infinity (we treat s as being a uniform random variable in $\{-,+\}^n$). Therefore, the vector $(p_0^s, p_1^s, p_2^s, p_3^s, p_4^s)$ converges almost surely to one of the following vectors: $(1,0,0,0,0)$, $(0,1,0,0,0)$, $(0,0,1,0,0)$, $(0,0,0,1,0)$ or $(0,0,0,0,1)$. In particular, p_k^s converges almost surely to 0 or 1.

Since $p_k \leq p_{k'}$ then $p_k^s \leq p_{k'}^s$ for any s, and so p_k^s cannot converge to 1 because otherwise the limit of $p_{k'}^s$ would also be equal to 1, which is not possible since none of the 5 possible vectors contain two ones. We conclude that p_k^s converges almost surely to 0, which means that $p_k^{(n)}$ (the average of p_k^s on all possible $s \in \{-,+\}^n$) converges to 0. Therefore, $p_k^{(\infty)} = 0$. ■

Proposition 6. *If $p_3 \leq \max\{p_1, p_2\}$, then we have maximal loss in the dominant face.*

Proof: If $p_3 \leq \max\{p_1, p_2\}$, then by the previous lemma we have $p_3^{(\infty)} = 0$. Therefore, we have maximal loss in the dominant face (see *remark 3*). ■

Corollary 2. *If we do not have maximal loss in the dominant face then the final capacity region (to which the capacity region is converging) must be symmetric.*

Proof: From the above proposition we conclude that $p_3 > \max\{p_1, p_2\}$ and from lemma 9 we conclude that $p_1^{(\infty)} = p_2^{(\infty)} = 0$. Thus, $I_1^{(\infty)} = I_2^{(\infty)} = p_3^{(\infty)} + p_4^{(\infty)}$ and the final capacity region is symmetric. In particular, it contains the "equal-rates" rate vector. ■

Conjecture 2. *The condition in* proposition 9 *is necessary for having maximal loss in the dominant face. i.e. if $p_3 > \max\{p_1, p_2\}$, then we do not have maximal loss in the dominant face.*

IX. Conclusion

In this paper, we have used quasi-group operations to construct capacity achieving polar codes for arbitrary DMCs with a probability of error that is less than $o(2^{-N^{1/2-\epsilon}})$, where N is the block length. This result allowed us to construct polar codes for arbitrary MACs by using an appropriate Abelian group operation.

It was shown in this paper that being a quasi-group is a sufficient property for an operation to ensure polarization if it was used in the construction of polar codes. The determination of a more general property that is both necessary and sufficient remains an open problem.

In the case of MACs, it was shown in this paper that while the symmetric sum capacity is achievable by polar codes, we may lose some rate vectors from the capacity region upon polarization. We have studied this loss in the case where the channel is a combination of linear channels, and we derived a characterization of non-losing channels in this special case. We have also derived a sufficient condition for having maximal loss in the dominant face in the capacity region in the case of binary input 2-user MAC.

It is possible to achieve the whole capacity region of any MAC by applying time sharing and polar coding. An important question, which remains open, is whether it is possible to find a coding scheme, based only on polar codes and Arikan-like constructions, in which all the symmetric capacity region is achievable.

References

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3051 –3073, 2009.

[2] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 28 2009.

[3] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Information Theory Workshop, 2009. ITW 2009. IEEE*, 2009, pp. 144 –148.

[4] E. Sasoglu, "Polar codes for discrete alphabets," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, 2012, pp. 2137–2141.

[5] W. Park and A. Barg, "Polar codes for q-ary channels,," *Information Theory, IEEE Transactions on*, vol. 59, no. 2, pp. 955–969, 2013.

[6] A. Sahebi and S. Pradhan, "Multilevel polarization of polar codes over arbitrary discrete memoryless channels," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, 2011, pp. 1718–1725.

[7] E. Şaşoğlu, E. Telatar, and E. Yeh, "Polar codes for the two-user multiple-access channel," *CoRR*, vol. abs/1006.4255, 2010. [Online]. Available: http://arxiv.org/abs/1006.4255

[8] E. Abbe and E. Telatar, "Polar codes for the -user multiple access channel," *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 5437 –5448, aug. 2012.

[9] R. Nasser, "Polar codes for the m-user multiple access channels," *CoRR*, vol. abs/1112.1770, 2011.
[10] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006.
[11] E. Sasoglu, "Polar Coding Theorems for Discrete Systems," Ph.D. dissertation, IC, Lausanne, 2011. [Online]. Available: http://library.epfl.ch/theses/?nr=5219